

GEOGRAPHIC
INFORMATION SYSTEMS
AND LIBRARIES:
PATRONS, MAPS,
AND SPATIAL INFORMATION

Edited by
LINDA C. SMITH
MYKE GLUCK

UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA - CHAMPAIGN
BOOKSTACKS

**GEOGRAPHIC
INFORMATION SYSTEMS
AND LIBRARIES:
PATRONS, MAPS,
AND SPATIAL INFORMATION**

**PAPERS PRESENTED AT THE
1995 CLINIC ON LIBRARY APPLICATIONS OF DATA PROCESSING
APRIL 10-12, 1995
GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN**

**CLINIC ON LIBRARY APPLICATIONS
OF DATA PROCESSING: 1995**

GEOGRAPHIC
INFORMATION SYSTEMS
AND LIBRARIES:

PATRONS, MAPS,
AND SPATIAL INFORMATION

Edited by
LINDA C. SMITH
MYKE GLUCK

GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
1996

© 1996 by The Board of Trustees of the University of Illinois
ISBN 0-87845-097-1 ISSN 0069-4789

Printed in the United States of America
on acid-free paper

1995
cop.3

Contents

Introduction

Linda C. Smith	
Myke Gluck	1

Mapping Technology in Transition

Mark Monmonier	5
----------------	---

Cataloging Planetospatial Data in Digital Form: Old Wine, New Bottles—New Wine, Old Bottles

Mary Lynette Larsgaard	17
------------------------	----

Finding and Accessing Spatial Data in the National Spatial Data Infrastructure

Michael Domaratz	31
------------------	----

The Government Information Locator Service: Discovering, Identifying, and Accessing Spatial Data

William E. Moen	41
-----------------	----

Geographic Information Systems and Digital Libraries: Issues of Size and Scalability

Barbara P. Battenfield	69
------------------------	----

Geographic Information Retrieval and Spatial Browsing

Ray R. Larson	81
---------------	----

Spatial Access to, and Display of, Global Change Data: Avenues for Libraries

Linda L. Hill	125
---------------	-----

Geospatial Information Needs of the General Public: Text, Maps, and Users' Tasks

Myke Gluck	151
------------	-----

Emergency Preparedness and Response Challenges for Special Libraries

Robert Lee Chartrand	173
----------------------	-----

Using Geographic Information Systems for Estimating and Profiling Geographic Library Market Areas

Christie Koontz	181
-----------------	-----

Implementing Geographic Information Systems in the Public Library Arena	
Dean K. Jue	195
The St. Louis Public Library's Electronic Atlas: A Successful GIS Application in the Public Library Environment	
Anne Watts	213
Digital Spatial Data of Illinois on CD-ROM, Department of Energy and Natural Resources	
Mark Joselyn Sheryl G. Oliver	221
Contributors	225
Index	231

LINDA C. SMITH & MYKE GLUCK

Introduction

Electronic technologies, including geographic information systems (GIS), are creating new ways of meeting the needs of library users for spatial and cartographic information. The 32nd Annual Clinic on Library Applications of Data Processing, held at the Beckman Institute on the campus of the University of Illinois at Urbana-Champaign (UIUC) on 2-4 April 1995, addressed the theme of "GIS and Libraries: Patrons, Maps, and Spatial Information." Current interest in this topic is evident in the publication of several special issues of journals shortly after the clinic took place. "Making GIS a Part of Library Service" (Lutz, 1995), "Geographic Information Systems (GISs) and Academic Libraries" (Hernon, 1995), and "Global Change and the Role of Libraries" (Rand, 1995) can be consulted for papers that complement and supplement the conference papers compiled in this volume. In particular, Longstreth (1995) identifies important information sources on GIS in his discussion of GIS collection development, staffing, and training.

GIS AND LIBRARIES

In his keynote address, Mark Monmonier presents a retrospective look at his book, *Technological Transition in Cartography*, published in 1985. The book examined the future of cartographic technology and the role of policy in the collection, dissemination, and use of spatial information. Monmonier hoped that readers would develop an understanding of the idea that the product of cartography is the information, not the image. His clinic paper briefly surveys each of the book's seven chapters and notes that many of the issues addressed remain pressing subjects today.

Four papers in this volume consider issues involved in describing spatial data sets and in organizing and providing access to digital libraries of spatial data. Mary Lynette Larsgaard examines the applicability of traditional library cataloging methods—e.g., Anglo-American Cataloguing Rules, USMARC—to catalog what she terms "planetospatial data" in digital form. Drawing on experience gained in cataloging resources for Project Alexandria, one of six projects funded by the National Science Foundation's Digital Library Initiative, she identifies several problem areas that need to be resolved in order to produce good catalog records. Michael Domaratz (in a paper based on a transcript of his presentation at the clinic) reviews the work of the Federal Geographic Data Committee in developing Content

Standards for Digital Geospatial Metadata (data about data) to provide support for a National Spatial Data Infrastructure. He also discusses the emergence of the National Geospatial Data Clearinghouse as a distributed, electronically connected, network of geospatial data producers, managers, and users. The intent of the clearinghouse is to allow users to determine what geospatial data exist, find the data they need, evaluate the usefulness of the data for their applications, and obtain or order the data as economically as possible. William E. Moen provides an overview of the Government Information Locator Service (GILS), a new federal initiative to assist the public in discovering, identifying, and locating government information. The basic components of GILS are: structured records with standardized data elements that describe and provide access information to federal information resources; agency-based information servers hosting these records; client software to support information retrieval from servers; and ANSI/NISO Z39.50 as the communications protocol between clients and servers. Because spatial data are an important category of government information resources, GILS may facilitate the identification and use of spatial data. Barbara P. Battenfield observes that as the size of a digital library increases, challenges for data organization and collection maintenance will also increase. She describes how models that are used in the physical and social sciences to predict growth and changes in size of particular phenomena can be applied to the growth of digital libraries. Allometric principles can be used to estimate the scale at which existing procedures will fail and new procedures must be implemented to handle further growth.

Ray R. Larson and Linda L. Hill explore issues in system design for enhanced access to geographic information and spatial data. Larson describes the characteristics of geographic information retrieval and spatial querying, examines the advantages of spatial browsing as a method of presenting a variety of georeferenced information in a coherent framework, and analyzes the feasibility of automatic indexing of geographic information embedded in text. Hill considers geospatial retrieval systems within the framework of the U.S. Global Change Data and Information System. She presents characteristics of an ideal geospatial retrieval system and describes five examples of the types of spatial retrieval available today. Myke Gluck demonstrates the value of experimental research in addressing a series of related questions: What are the geospatial information needs of the general public? What are the different formats and tasks for geospatial information suggested by the public? What formats for geospatial information are most useful under differing task situations? and What role may the public library play in assisting to resolve geospatial information needs for the public? Findings of Gluck's studies are helpful in suggesting ways that libraries and system designers can improve access to geospatial information.

Robert Lee Chartrand and Christie Koontz provide examples of applications of GIS. Chartrand emphasizes the importance of geographic-oriented information for emergency preparedness and response and argues that special libraries in particular should be prepared to fill this need. Koontz demonstrates the applicability of GIS to library market analysis by graphically estimating geographic boundaries and analyzing socioeconomic characteristics within prescribed markets. Examples are drawn from an analysis of branch location and populations served by the Evansville-Vanderburgh County public library system.

The last three papers consider issues in implementing GIS. Dean K. Jue draws on a survey of libraries that have introduced GIS, in order to identify factors associated with successful implementations. He also presents a decision flowchart to help public librarians evaluate the type of GIS services that could be provided in any given library environment. Anne Watts offers a case study of a successful GIS application at the St. Louis Public Library: an electronic atlas of 1990 census tract maps and data for St. Louis City and County. She describes the system design and how it has been used as a public workstation. Watts notes several factors contributing to the success of the project: well-defined and limited system, collaboration between the library and outside experts in its development, an internal organizational champion, and approaching the project as a natural extension of the library's information services. Mark Joselyn (in a paper presented at the clinic by Sheryl G. Oliver, GIS Manager, Illinois Department of Energy and Natural Resources) provides an example of a state government initiative to gather and make available spatial digital data. Data sets contained on the Illinois CD-ROM were primarily constructed by divisions of the Illinois Department of Energy and Natural Resources, with others derived from U.S. Geological Survey and census files. The CD-ROM was distributed free of charge to government agencies, schools, and libraries.

OTHER COMPONENTS OF THE CLINIC

In addition to the papers presented in this volume, the clinic included a talk by Brent Allison, head of the John R. Borchert Map Library at the University of Minnesota, on "GIS in Academic Libraries" (see Allison [1995] for a recently published account of Minnesota's Automated Cartographic Information Center). A presentation on "Local Initiatives in GIS" by Douglas Johnston of UIUC and R. Christopher Schroeder of AEC Centrec Consulting Group, Savoy, Illinois, highlighted activities of CCNet in Champaign County, Illinois, and focused on GIS issues for agriculture. The clinic began with two preconference workshops covering an overview of GIS concepts (Mark P. Armstrong, University of Iowa) and spatial analysis in the social sciences (Gerard Rushton, University of Iowa). Demonstrations drew on individuals

associated with the U.S. Army Construction Engineering Research Laboratories and UIUC to provide examples of GIS systems and applications. The editors gratefully acknowledge the contributions of all these individuals to the success of the clinic.

LINDA C. SMITH
MYKE GLUCK
Editors

REFERENCES

- Allison, B. (1995). University of Minnesota: Remote access and the Internet. *Journal of Academic Librarianship*, 21(4), 283-287.
- Hernon, P. (Ed.). (1995). Geographic Information Systems (GISs) and academic libraries. [Special issue]. *Journal of Academic Librarianship*, 21(4), 231-296.
- Longstreth, K. (1995). GIS collection development, staffing, and training. *Journal of Academic Librarianship*, 21(4), 267-274.
- Lutz, M. (Ed.). (1995). Making GIS a part of library service [Special issue]. *Information Technology and Libraries*, 14(2), 77-122.
- Monmonier, M. S. (1985). *Technological transition in cartography*. Madison, WI: University of Wisconsin Press.
- Rand, R. Y. (Ed.). (1995). Global change and the role of libraries [Special issue]. *Library Hi Tech*, 12(1-2), 7-84.

MARK MONMONIER

Mapping Technology in Transition

This presentation is a light-hearted mildly self-critical recapitulation of a 1985 book-length essay on the evolution and future of cartographic technology, which emphasizes the significant role of public policy in the collection, dissemination, display, and use of geographic information.

INTRODUCTION

However self-serving this talk might appear, it should be rewardingly informative—or at least informatively entertaining. The subject of my talk is a book—a book most publishers would consider a commercial failure. In the decade since its publication in 1985, *Technological Transition in Cartography* has sold fewer than 1,200 copies in addition to the roughly 200 copies the publisher seems to have given away. Even so, the University of Wisconsin Press still offers the book on its backlist, a service few trade publishers would dare provide.

Truth be told, Wisconsin was not my first choice for a publisher. I had offered the book to the University of Chicago Press, which turned it down after the first reader objected strongly to what she interpreted (quite incorrectly) as my promilitary viewpoint. My sin, it seems, was an unscholarly fascination with cruise missiles, those low-flying pilotless planes guided to their targets by digital maps, terrain scanners, and global positioning systems. Even now I find this technology not only fascinating but—if a war is unavoidable and the technology works like it should—ininitely preferable to napalm, carpet bombing, and a nuclear holocaust. University press boards react negatively to such criticism, though, and I agreed with Chicago's acquisitions editor that I should try elsewhere (ironically, my long-standing relationship with Chicago began several years later, when I approached the press with the manuscript for *Maps with the News* (Monmonier, 1989), which the Johns Hopkins Press had rejected after an even more vitriolic first reader objected that my way was not his way). Fortunately, my good friend David Woodward introduced me to the director of the University of Wisconsin Press. A pair of more open-minded readers agreed I had something worth saying, and the rest is cartographic history.

Technological Transition in Cartography was not only a fun book to write but a convenient way to reach cartography students, too many of whom seemed very narrowly focused on the technology of drawing maps with

pen and ink, scribing tools, or pen plotters—manual skills with a very short half-life. I wanted to promote a broader view of mapping as a field in which technology was evolving rapidly and in which the principal product was geographic information, not the printable image. And I wanted to show them that dramatic changes readily apparent in the early 1980s would prove as deep seated and far-reaching as such earlier cartographic revolutions as printing and aerial photogrammetry.

ORGANIZATION AND CONTENT

The book was divided into seven chapters: an introduction, a summary and conclusions, and individual essays on location and navigation, boundaries and surveys, aerial reconnaissance and land-cover mapping, decision support systems, and map publishing and the digital map. In the introduction, a historical comparison of map copyrights afforded a brief examination of change in the nature of map products, which I related as well to the mapmaker's ability to adapt a wide variety of innovations, institutional as well as technological. The final chapter speculated on various ways mapping seemed likely to change as a result of the ongoing revolution, which I called the "electronic transition."

Although concrete examples were important, my approach was far more nomothetic than idiographic. Each chapter examined a variety of technological changes and speculated on the impact of recent innovations. Chapter two, which examined advances in location and navigation, attempted to place the cruise missile and global positioning system in perspective with a concise examination of earlier navigation aids, such as the Marshall Island stick charts, the Mercator projection, the chronometer, and sonar. The key conclusion in the chapter, though, was the effect on civilian mapping and navigation of "trickle down" from military and space. In no way an endorsement of the "trickle down" claimed by Reagan appointees advocating supply-side economics and massive tax cuts, my points were that "the perceived need for ever better national security will remain the principal impetus for major new developments in mapping and map use. [After all,] a good defense system requires accurate geographic intelligence and accurate navigation. Only the comparatively lavish appropriations for defense are likely to sustain the current rate of development in digital cartography. Mapping thrives on war and threats of war" (Monmonier, 1985, p. 45). The implications of all of this for the general public seemed clear: "Before [the year] 2000, [the map user] might well push a button on a small, portable pocket navigator and read out his coordinates accurate to several decimal places" (p. 45). Although I had little to say about automatic vehicle navigation systems like the Etak Navigator, it seemed clear that: "Knowing where you are and how to get to where you want to go will no longer be challenging problems" (p. 45).

Chapter three, on boundaries and surveys, looked at advances in geodesy and land survey. Early illustrations describing the plane table, leveling rod, and surveyor's transit led to discussions of electronic distance measurement, triangulation networks, aerotriangulation, stereophotogrammetry, orthophotomapping, and inertial positioning systems. My concluding comment, though, identified similarities between geodetic control systems and the highway network: "Both facilities are geographically extensive; both require careful coordination and continual maintenance. Because direct user charges are troublesome as well as costly to assess, both networks are public-sector obligations, supported largely from general [tax] revenues" (Monmonier, 1985, p. 74). I drew other parallels as well, including recognition that even though a "surveying and mapping lobby" consisting of users, contractors, and equipment manufacturers is smaller and less powerful than the highway lobby, "both mapping and road building are influenced at least as much by political considerations as by progress in science and engineering" (p. 75).

The fourth chapter, titled Aerial Reconnaissance and Land Cover Inventories, also addressed topographic mapping, perhaps the biggest beneficiary of advances in aerial survey. I began by looking at the county atlases sold by subscription from the 1820s onward—the comparatively crude private-sector forerunner of the public-sector topographic map series. The U.S. Geological Survey, founded in 1879, initiated a topographic survey in 1882 with 15-minute quadrangles mapped at a scale of 1:62,500 and 30-minute quadrangles mapped at 1:125,000. A case study of Pennsylvania noted that state support of federal mapping, called "cooperation," had an important leveraging effect: the Commonwealth was mapped much more rapidly than the nation as a whole because the Geological Survey was responsive to states that contributed to USGS mapping activities within their borders. Ironically, when the Geological Survey appraised the quality of its surveys in 1946, the most recently mapped sections of the Commonwealth—the relatively remote counties of the "Northern Tier"—had the more accurate, more up-to-date maps. Also examined were the temporal trends in the growth of the 7.5-minute series of more detailed 1:24,000 maps. Because of aerial photogrammetry, spatial coverage of the 7.5-minute series advanced much more rapidly than for the 15-minute series, begun in an era of plane-table surveys and topographic field sketching.

The chapter also examined advances in aerial survey, including color-infrared imagery developed by the military for camouflage detection and electronic sensing systems, including active side-looking airborne radar (SLAR) systems as well as passive multispectral scanners. Image processing techniques developed with satellite remote sensing include: the simple parallelepiped classifier and more sophisticated classification algorithms, resampling to remove geometric distortions, edge enhancement to sharpen detail, and smoothing enhancement to remove the effects of "spikes" of

high or low brightness. Remote sensing not only revolutionized the mapping of land cover and land use but held out the possibility of more accurate choropleth maps of population density computed by dividing county populations by the amount of residential land.

The chapter concluded with a brief discussion of mapping policy. More so than earlier mapping techniques, remote sensing raised issues of tradeoffs and privatization. At one level, civilian-oriented remote sensing programs like Landsat could be optimized for geologists or agriculturists—that is, to serve scientists concerned with geologic structure and petroleum exploration in contrast to scientists focused on vegetation and the world food supply. And of course there was the inherent conflict between civilians eager for the detailed imagery that was technologically possible and defense experts concerned that providing high-resolution imagery for civilian applications was a threat to national security even though our main opponent (the USSR) most certainly had excellent satellite spy systems. Amazing what a little competition from the French SPOT system did to that argument—not to mention the availability in the early 1990s of high-resolution satellite imagery from the former Soviet Union.

Remote sensing raised many other policy issues including the needs of academic researchers and other users less able to pay and the objections of Third World nations who charged that unauthorized sales of imagery of their national territory was a breach of privacy rights and international etiquette. Even so, “a tactfully managed satellite sensing program can be an effective instrument of international cooperation and regional development” (Monmonier, 1985, p. 109).

My fifth chapter, on decision support systems, began by looking at censuses and surveys for collecting demographic and socioeconomic data and moved quickly to the DIME (dual independent map encoding) data structures used to support address matching and computer mapping for the 1970 census. The chapter then briefly examined the history of computer mapping technology, which had advanced in two decades from comparatively crude line-printer maps to a variety of higher-resolution hard-copy displays based on the pen plotter, to interactive cathode ray tube displays with pointing devices and “user-friendly” menus. Of particular interest were plasma panels and other comparatively thin wall-panel displays as well as computer-controlled holograms.

The chapter turned from display to data in a section on National Atlases and Data Banks. After a concise look at the history of national atlases as instruments of planning, discovery, and education, I proposed that “modern telecommunications and data base technology will radically alter the organization and operation of national atlas programs,” with the paper atlas “secondary to the digital cartographic data base from which it will be derived” (Monmonier, 1985, p. 136). The forecast included data

transmitted by “wire or airwaves,” more timely analyses in home or office, an improved cost-benefit ratio, and a fuller role for the private sector. Information policy would become a significant policy issue for cartographers and geographers, with free or low-cost public access a key issue.

Equally troublesome was the need for coordination, as much to thwart the easy addition of misleading poorly documented information as to provide open access to users. “Ready availability does not guarantee quality,” I noted (Monmonier, 1985, p. 137). “If poorly controlled, an easily accessible data base may do more harm than slower, more traditional mapping methods” (p. 137). Also important was coordination with state and local agencies, with the possible addition of regional atlases and electronic databases. States would want their own electronic atlas systems, and state and federal systems would need to learn to talk to each other.

The next section addressed turnkey systems, mapping software, and usability. Of particular concern were the default options by which software developers gave naïve users a success experience by imposing standardized design decisions, often of dubious value. Nowhere was this more evident than in statistical mapping software that instantly generated colored five-category choropleth maps based on an equal-intervals classification. “Cartography as a profession,” I noted, “should be concerned with the inventory and comparative evaluation of mapping software, with the clear and comprehensive documentation of program goals and system operation, and with the proper training of program users” (Monmonier, 1985, p. 142). Although cartographic journals have been rigorously reviewing software for nearly a decade, the profession today seems as powerless against stupidly designed software as it was against the flagrant misuse of the Mercator projection.

The final section, on Standards, Cooperation, and Shared Benefits, offers an optimistic note. In a broad plea for efficient documentation, compatibility of data and software, and widely recognized data exchange formats, I called for institutional guarantees that data are reliable, complete, and valid. Although standards offered some hope, “standardization can be both good and bad” (Monmonier, 1985, p. 143). In particular, I warned of a “hasty, politically expedient imposition of standards [that] might greatly lower the benefits of coordination.” Like so many things, the effectiveness of standardization depends on how well it is done.

Chapter six, titled Map Publishing and the Digital Map, began with a brief examination of two earlier innovations with broad effects on cartography—namely, printing and photographic engraving. Of particular interest were unique cartographic adaptations such as wax engraving, which linked letterpress printing with map drafting through a nineteenth-century process of electroplating. An efficient method for

rendering curved type as well as standardized point and line symbols, wax engraving gave railroad and textbook maps a distinctive look.

A more recent cartographic innovation, negative scribing with plastic film and "peelcoats" reached its heyday in the 1970s. Useful for producing cartographic layers known as "feature separations" and "color separations," this technology dominated advanced classes in map design and reproduction. And for a decade or so before the advent of the high-resolution imagesetter, digital map-production technology focused on automating the manual aspects of negative scribing. By the early 1980s, though, it was apparent that scanners would capture existing hardcopy cartographic images, while sophisticated commercial mapmakers created (and stored) new ones electronically.

In a section on Computer Memories for the Digital Map, I looked at electronic storage media ranging from cassette tapes and floppy discs to semiconductor memory and computer-searchable laserdiscs (videodiscs)—technology able to "place a sizable map collection in millions of homes" (Monmonier, 1985, p. 163). And in a section titled Glass Threads, Networks, and Videotex, I examined an emerging technology that promised maps a wider role in such information services as "city information, boating/fishing information, weather forecasts, real estate sales, and job searching" (p. 178). Even so, cartographic experiences in the development of printing and photography suggested this role would be secondary and passive rather than primary and active. "While not a prime innovator," I pointed out, "the map maker must be an efficient and innovative adaptor" (p. 178).

Chapter six concluded with a short essay on Maps as Software and a call for less emphasis on the paper map. "The Electronic Transition seems destined to progress to a stage at which maps are seldom composed on paper or similar [hard-copy] graphic media.... With time," I argued, "almost all cartographic data will be captured electronically, at their source, through photogrammetry or satellite remote sensing, or from computer systems for processing census results and other administrative data." Consequently, "a prime casualty of cartography's Electronic Transition will be the attitude that the map is a printed product.... [And although] it might be too soon to forecast the demise of the paper map, paper clearly [would] become less important as a storage medium and vehicle for geographic information.... [In slightly different words,] the paper map might not be dead, but the digital map will greatly alter where, how, and when paper maps are printed" (Monmonier, 1985, pp. 178-179).

THE BOOK'S CONCLUSIONS

The final chapter of the book was an opportunity not only to collect various observations from the earlier chapters but also to speculate more broadly about the future of maps, mapmaking, and mapping policy. I began with an attack on cartographic Darwinism: the naïve assumption "that

maps will be better—more accurate, more timely, more accessible, more aesthetic, more tailored to user needs—simply as a result of high technology.” More so than ever before in the history of cartography, the greatest challenge was institutional—the need for effective management and organization. “The potential for better maps has never been greater,” it was clear, “yet neither has the threat of dismal, expensive, embarrassing failures” (Monmonier, 1985, p. 181).

The key issues involved public policy, not technology. Prime concerns in the early 1980s (and today as well) were privacy, public access, and cost recovery. Electronic linking of heretofore segregated records and high-resolution satellite sensing systems opened new avenues for questionable snooping on citizens by government and private individuals alike. Yet these tools might prove at least equally adept in legitimate attacks on crime and environmental pollution. Even so, who should have access to these and less sensitive data and at what price? Because electronic geographic data can be costly to develop, private developers need copyright protection. In the early 1980s, though, politicians and federal bureaucrats were beginning to think about a variety of radical strategies—radical for the United States, at least—including copyright protection for federal databases and privatization of heretofore public data-collection activities like the National Weather Service. Driving much of this thinking was recognition of the enormous value of geographic data to corporations that could well afford to pay for much more than the marginal cost of printing and distribution. Yet equally apparent was the difficulty of establishing a schedule of charges that provided fuller cost recovery yet accommodated the needs of academic researchers and other citizens with much shallower pockets. And turning mapping over to a for-profit corporation might well lead to neglect of areas with limited sales potential. To safeguard the needs of scholars and hobbyists as well as to shield national mapping programs from “blind allegiance to cartographic Nielsen ratings,” I proposed (slightly tongue-in-cheek) a National Endowment for Cartography.

Coordination of mapping was an important policy need, which federal officials traditionally dealt with through interagency committees. Was it not time, I asked, for a major cabinet-level realignment, such as the consolidation of all mapping activities in a new Department of Natural Resources? Such proposals were by no means new, though, and I was hardly optimistic, especially about the likelihood of stronger links between topographic mapping on the one hand and demographic and economic censuses and surveys on the other. Although the military and natural resource agencies in the departments of Interior and Agriculture had a long history of communication and cooperation, obvious ties between mapping policy and statistical policy were markedly weaker.

Another concern was the growing role of private enterprise, which even without privatization would have substantial impacts on mapping.

Especially powerful was the multinational corporation, with needs to explore and integrate its operations “nationally and internationally without regard to political boundaries” (Monmonier, 1985, p. 188). Although largely an afterthought, the often subtle cartographic role of multinational corporations seems an intriguing topic for at least one doctoral dissertation on the history of late twentieth-century cartography.

To professional colleagues, my next section on Security and Digital Data must have seemed like science fiction if not outright hallucination. A clear drawback of electronic storage was the dual threat of “nuclear attack and the malevolent prankster,” I wrote (Monmonier, 1985, p. 189). Like other forms of electronic data, digital cartographic information was vulnerable to electromagnetic pulse (EMP), whereby a high-voltage wave produced by a high-altitude nuclear explosion could shut down the electric power grid, disable telecommunications, and wipe out fragile circuitry and electronic storage media. Equally problematic, albeit in a wholly different way, was the capricious hacker who might move Cleveland into Lake Erie, delete important obstructions from aeronavigation charts, or enlarge his or her own lot at the expense of a neighbor. Like other electronic information, cartographic data must be defended against tampering and EMP.

A section on Preservation and the Historic Record addressed another challenge of cartography’s electronic transition. Despite the vulnerability of paper to heat and time, paper maps are more easily collected, cataloged, and preserved than their electronic counterparts. Because hard-copy electronic storage is far from permanent, digital data must be recopied periodically. Moreover, preservation of electronic maps also requires preservation of display software and associated operating systems. Especially troublesome is the geographic database that is updated on a more or less continuous basis. Although edits and other changes might be archived so that any desired cartographic snapshot can be reconstructed, the map author or database administrator must make a conscious effort to preserve the cartographic record.

But what exactly does “cartographic record” mean in the digital era? I am not aware that any library or archives is systematically preserving late twentieth-century electronic cartography. Yet the challenge is enormous because an adequate historical record would include not only maps, data, software, and other artifacts but also information on how people, companies, and governments are using cartographic data. Future historians of cartography can benefit greatly from efforts to gather oral histories, institutional records, reports of participant observation, and recordings of interactive map analysis sessions. The history of cartography, unfortunately, seems too heavily committed to an antiquarian focus on the field’s more distant past. And even unsystematic collectors short-sightedly avoid contemporary maps.

My final section on Humanistic Challenges addressed design, aesthetics, and the growing number of do-it-yourself mapmakers empowered by software and digital data. Unencumbered by either intelligent software or cartographic training, untrained users can be viewed as a threat to both themselves and other equally naïve users of their creations. This view, of course, is sometimes attacked in these postmodern times as the elitist hand-wringing of “the professional” with an economic and ideological stake in cartographic education—a charge not without merit. Even so, perhaps the greatest scholarly challenge of cartography’s electronic transition was the apparent democratization of mapmaking that not only spawned millions of poorly designed maps but gave hundreds of thousands (at least) of novice mapmakers the opportunity to create potentially impressive cartographic artwork with crisp type and authoritative symbols, and to easily integrate maps with writing. Although I didn’t emphasize this point, cartography’s electronic transition offers the potential at least of a philosophical-cognitive revolution in how we think about and describe spatial relationships.

CRITICAL REVIEWS

Before closing, I want to address two fairly obvious, but nonetheless important, questions: How was the book received? And what do I wish I had treated differently?

First, the book’s impact, which can be assessed in part through an examination of published reviews. Some of the two hundred free copies, struck responsive chords with book review editors, and critiques appeared in most (if not all) appropriate journals in geography and cartography, including *The American Cartographer*, the *Annals of the Association of American Geographers*, the *Bulletin of the Society of University Cartographers*, *The Canadian Geographer*, the *Cartographic Journal*, *Cartographica*, *Computers and the Social Sciences*, *Environment and Planning A*, *The Geographical Review*, *Geography*, *The Journal of Geography*, *The Photogrammetric Record*, *The Professional Geographer*, and *Technology and Culture*. More gratifying, most reviewers seemed pleased, although some were more enthusiastic than others, and even enthusiastic ones were a bit picky at times.

I kept track of reviewers’ assessments by extracting their most positive and most negative comments as well as by rating each review A, B, C, or D (all very scientific, of course). Fortunately, there was no need for an F category—in the mid-1980s, that needlessly obtuse ritual whining known as postmodern critique had little interest in maps and geography. By my count (and you’ll have to take my word for this), the book earned six As, twelve Bs, one C, and one D—a skewed distribution not unlike the grades I award in upper-division courses in advanced map design or geographic analysis.

The most enthusiastic reviewers not only acknowledged the book's clarity and timeliness but agreed that maps were due for profound change, and that cartographic scholars had seriously neglected the importance of policy. As an illustration of an "A" review, Michael Blakemore (1986), writing in *Environment and Planning A*, opined: "The joy of this book is that it provides a scholarly synthesis of the past as well as posing a wide range of questions about future issues. . . . It becomes increasingly clear from this excellent book that digital cartography throws up issues at such a rate that society feels it difficult even to react, let alone think the issues through logically" (p. 283).

The "B" reviews typically balanced positive comments with doubts or reservations. For example, Alan MacEachren (1986), writing in *The American Cartographer*, enthusiastically stated: "This book is unique in both its topic and perspective . . . this is a book that belongs in every cartographer's library. It contains opinions on cartography's future that deserve consideration by everyone involved with making, studying, and using maps." Yet he was critical of its content and focus: "As an introductory text, the book is rather narrow in scope. . . . The emphasis on national mapping policy and programs, while justified, so dominates the book that a beginning student might assume that all significant cartographic developments take place in the federal government."

The view that book-length essays on cartography are perforce textbooks accounts for a number of less than laudatory comments in the two "C" and "D" reviews. The "C" review—by a graduate student (Noronha, 1986) whose advisor (ironically) penned what I consider an "A" review (Goodchild, 1987)—had little positive to say beyond noting that "the language is clear" and that the "wealth of factual detail tucked between the lines . . . makes for a comprehensive historical perspective" (p. 309). The reviewer was deeply disappointed by the omission of such theoretical concepts as "smoothing, Fourier analysis, and fractal geometry" and by the dismal green dust jacket, the aesthetically mediocre graphics, and the lack of color illustrations. And the "D" review (Slocum, 1986), while conceding that: "Generally, this book does meet its intended goal," and concludes that "it is unlikely that [the book] will function well at the introductory cartography level because it is too broad in scope. It is more appropriate for a history of cartography course, or possibly a computer cartography seminar. This book might also be useful in a history of science course because it discusses a number of technological advances ancillary to cartography" (p. 38). Although I agree with, and even applaud, this assessment, I obviously should have emphasized more strongly my intent not to write an introductory textbook.

REGRETS AND LOOKING AHEAD

What else would I have emphasized? Two things, principally: an increased facility for integrating graphics with written text and the advent of dynamic experiential maps able to free users from ossifying single-map views.

I addressed integration more satisfactorily four years later in another university press book, *Maps with the News: The Development of American Journalistic Cartography*. The final chapter concluded with the observation that personal computers and mapping software held the promise of richer geographic descriptions, interpretations, and explanations by journalists and writers able to link both media with ease and think graphically as well as verbally (Monmonier, 1989, pp. 246-47). This consequence of cartography's electronic transition is clearly deeper and more far-reaching than my earlier forecast of a "de-massified' cartography" (Monmonier, 1985, p. 172).

Equally significant is the likelihood of experiential maps. Defined as dynamic displays with which the viewer can readily interact, experiential maps are much more than interactive mapping software: guided by scripts and user profiles, an experiential map can explain a process or act out a temporal event with terminology, symbols, sequences, and geographic foci adapted to the viewer's interests and experience. At its simplest level, the experiential map might reveal geographic nuances hidden by a traditional static map. At more advanced levels, it can orient a viewer to a new atlas or electronic database, help the user comprehend complex concepts in physical or human geography, or assure that an analyst will not overlook a potentially meaningful spatial pattern. Although my 1985 essay forecast an increased use of both animation and interactive mapping (Monmonier, 1985, pp. 137-42, 163), more recent developments, such as graphic scripts (Monmonier, 1992), suggest profound changes in how maps can engage users in a search for knowledge and understanding.

REFERENCES

- Blakemore, M. (1986). [Review of the book *Technological transition in cartography*]. *Environment and Planning A*, 18, 283.
- Carter, J. R. (1986). [Review of the book *Technological transition in cartography*]. *The Geographical Review*, 76(1), 112-113.
- Gardiner, V. (1986). [Review of the book *Technological transition in cartography*]. *Bulletin of the Society of University Cartographers*, 20(1), 56-57.
- Goodchild, M. F. (1987). [Review of the book *Technological transition in cartography*]. *Cartographica*, 24(3), 98-100.
- Hollway, J. (1986). [Review of the book *Technological transition in cartography*]. *Geography*, 71(1), 92.

MONMONIER/MAPPING TECHNOLOGY IN TRANSITION

- Karrow, R. W., Jr. (1987). [Review of the book *Technological transition in cartography*]. *Technology and Culture*, 28(1), 181-183.
- Keller, C. P. (1986). [Review of the book *Technological transition in cartography*]. *The Canadian Geographer*, 30(4), 379.
- MacEachren, A. M. (1986). [Review of the book *Technological transition in cartography*]. *The American Cartographer*, 13, 276-277.
- Monmonier, M. (1985). *Technological transition in cartography*. Madison, WI: University of Wisconsin Press.
- Monmonier, M. (1989). *Maps with the news: The development of American journalistic cartography*. Chicago, IL: University of Chicago Press.
- Monmonier, M. (1992). Authoring graphic scripts: Experiences and principles. *Cartography and Geographic Information Systems*, 19(4), 247-260, 272.
- Noronha, V. T. (1986). [Review of the book *Technological transition in cartography*]. *The Professional Geographer*, 38(3), 308-309.
- Slocum, T. A. (1986). [Review of the book *Technological transition in cartography*]. *The Journal of Geography*, 85(1), 38.
- Snyder, J. P. (1987). [Review of the book *Technological transition in cartography*]. *Annals of the Association of American Geographers*, 77(1), 148-150.
- Walker, A. S. (1986). [Review of the book *Technological transition in cartography*]. *The Photogrammetric Record*, 12(67), 120-121.

MARY LYNETTE LARSGAARD

Cataloging Planetospatial Data in Digital Form: Old Wine, New Bottles—New Wine, Old Bottles

This discussion deals with using traditional library cataloging methods—e.g., Anglo-American Cataloguing Rules (AACR), USMARC—to catalog planetospatial data in digital form and the problem areas that have come to light.

INTRODUCTION

University research libraries are in some ways like modern medicine, most noticeably in their tendency toward specialization. For example, an ear-eye-nose-and-throat doctor may not be the person you would like to trust to take care of an ailment that happens south of the larynx. Similarly, in the approximately twenty years to 1988 that I have been occupied with (among other occupations) cataloging maps, I never once cataloged digital data (but I have an excuse—no digital data in the library with the exception of the online catalog) (distant sound of the gods laughing).

At the same time, it was obvious, even ten years ago, to many in map librarianship that our portion of Library Land was going digital. Thus the last five or so years have presented many learning experiences, especially since Project Alexandria started in October 1994. Project Alexandria—or, as it is also called, the Alexandria Digital Library—is one of six National Science Foundation-funded Digital Library Initiatives (DLI). Alexandria's goal is to provide online access to georeferenced information starting with planetospatial data. My role in Alexandria is multipartite, focusing on metadata, data set selection (both metadata and spatial data), and general information on library services and practices—what works, what does not, and what services we would like to provide to users. Very specifically, since early October of 1994, I have had primary responsibility for working with a computer engineer to design a metadata schema (called the Alexandria Metadata Schema). This metadata schema is a brief version, with several simplifications (e.g., no punctuation between fields) of USMARC, with special emphasis on the fields of the U.S. Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata. Mike Domaratz, a main architect of the later standard and another presenter at this clinic, will provide specific comments about that. Other than some allusions that reflect the work done for

Alexandria over the last several months, what is presented in this discussion is a snapshot of a work in progress as I decide how to deal with the moving target of digital data forms. Honesty compels me to admit that the Alexandria Metadata set is composed of a paltry 130 records, only about half of which are digital data (although all of the analog items have been scanned).

It is from this experience whence comes the subtitle 245\$b. “Old wine, new bottles” in the title of this presentation refers to information formerly presented in paper, and other generally eye-readable, versions now being presented in digital form—initially in somewhat literal transformations but increasingly in new ways. The “new wine, old bottles” refers to the tremendous fun we have fitting new formats into old cataloging rules, adding or changing the latter as new data to be accessed appears or as existing types develop new wrinkles. As for the adjective “planetospatial,” the rationale follows that the term “map” is obviously inadequate. How about the term “spatial data”? That encompasses too much since it includes measurement of any object in space (e.g., medical imaging). The terms “geospatial data” and “georeferenced information,” however, each begin with a word derived from the root “ge-”, that is, Earth, so data relating to other planets would not be covered. “Cartographic materials” comes close, but to many people it means maps only. For example, there is a feeling that the MARC Map format is to be used only for maps, which is not correct—it may be used (at least in theory) for any cartographic material. “Georeferenced information/data” does indeed nicely cover anything that can be given a latitude/longitude reference—“spatial data that pertain to a location on the Earth’s surface” (Farrell, 1994, p. 1). Thus it includes, for example, gazetteers, population statistics, histories of countries, and so on. So that seems to leave us with “planetospatial data,” a term suggested by a computer science graduate student.

The two learning experiences that are the basis of this presentation are those of my self-education in cataloging digital planetospatial data, and of explaining Map and Imagery Lab materials and library cataloging practices to computer science (and other science) faculty members and graduate students. The latter has been occasionally—but fortunately, seldom so—an embarrassing process, since once I looked into something—and sometimes even after a quick glance—I found practices that were difficult or impossible to defend. In some cases, these practices seem to be based on the problem on which library science has been based—as the name indicates—on a format (the book) instead of on a concept (information). It has also been my observation, as I read various manuals on cataloging digital data, that there are differing ideas and perhaps some confusion as to how to catalog that data.

On the other hand, there have been moments of sweet revenge, as when a few months into putting together a prototype, a computer engi-

neer informed me, in tones of amazement and indignation, that compiling metadata was extremely time consuming. You will be pleased to know—maybe—that the efforts we have made to ensure that users do not have an inkling as to what goes on before an item is ready for use have succeeded beyond our wildest expectations. I have also learned that much of what we take for granted in online cataloging—e.g., having all fields, whether repeatable or nonrepeatable, on one form; having as many full-text (as compared to varchar/limited-to-256-characters fields as are needed); having the computer software “know” that, for example, Calif. = California = CA is no mean code-writing/database-management-software feat.

It has been most helpful to read the writings of people in other disciplines turning their minds and their discipline’s habit of thought to cataloging (see especially Bretherton and Singley [1994] to see how a model of catalog access to information looks to two atmospheric scientists). It is ironic that information derived by cataloging had to be called something else—metadata—before noncatalogers dealt with it. This is similar to persons who claimed they could not type who, all of the sudden, learned how to type when computers came in, and the name of the work was described as “inputting.” The bright side of dealing with metadata for digitally generated data is the promise (which one fervently hopes will not evanesce) that software (perhaps the tools of spatial databases and analysis) may be used to extract metadata. It cannot happen too soon for me.

It seems inevitable, in this self-education process, that there are some areas where cataloging GIS and other planetospatial digital data has meant the need for some retooling and reworking—and perhaps some reinvention—of existing cataloging policies and procedures. It is extremely fortunate that the purpose of cataloging policies and practice is to ensure that users will be able to get at the data in ways that they consider the most logical. This means that catalogers using jesuitical reasoning to get the access points needed is perfectly acceptable. Yet, overall, much of what exists in current cataloging practice works quite well. The following text will reveal exceptions to this premise. As part of the work-in-progress aspect of this discussion, it will be obvious that there are no answers for all of these concerns, although there are routes of attack.

But first let us take a glance backward. In the “The more things change, the more they remain the same” category, Dodd’s (1982) remarks as to why machine-readable data files may be difficult to catalog:

1. currently lack desirable standards in bibliographic representation; AACR2 (Anglo-American Cataloguing Rules, 2d ed.) does not adequately define bibliographic elements as they apply to MRDF;
2. lack of internal user labels: poor documentation, lack of such traditionally used elements as a title page although a standard for a header does exist (ANSI X3.27);

3. production rather than publication is the rule;
4. data are easily and, in some cases, often changed, so assigning editions may be difficult;
5. dates of publication may be problematic because of 3 and 4 above; and
6. physical description is a problem area, since the items are so different from traditional library materials (pp. 35-37).

Thus the following list of the problem areas I have seen will come as little surprise:

1. physical description, especially 300\$a generally and specifically SMDs but also 300\$b and the way this field relates to file-characteristics area (256) and mathematical data (255);
2. production versus publication;
3. merging another standard—Content Standards for Digital Geospatial Metadata with USMARC;
4. multilevel description;
5. looking in a mirror looking at self holding a mirror—when one catalogs an analog item (e.g., a map), which is scanned at high resolution and when does one stop cataloging? Does one catalog the scanned file separately?
6. subject headings;
7. bounding coordinates; and
8. time—which one?

PHYSICAL DESCRIPTION

General Material Designations (GMDs), Specific Material Designations (SMDs), and (not surprisingly) the USMARC formats are a mixture of intellectual and physical formats, of content and form, of information and carriers of information, and of frequency. It is tempting to say that they were heuristically developed in order to deal with existing data, that they work, that format integration enables one to use any field needed, and to leave it at that. The fact remains that putting, for example, “map” in the same category as “slide” is not logical.

General Material Designations (GMDs) in AACR2 include

List 1—braille, cartographic material, computer file, graphic, manuscript, microform, motion picture, multimedia, music, object, sound recording, text, videorecording.

List 2—art original, art reproduction, braille, chart, computer file, diorama, filmstrip, flash card, game, globe, kit, manuscript, map, micro-

form, microscope slide, model, motion picture, music, picture, realia, slide, sound recording, technical drawing, text, toy, transparency, videorecording (Gorman & Winkler, 1988, p. 21).

USMARC formats include: book, serial, archives and manuscript control, machine-readable data files, maps, music, visual materials.

Specific Material Designations (SMDs) include: monographs in book form: no SMDs referred to as such—v., p., leaves, columns, broadside, sheet, portfolio (p. 72).

Serials designations include: Use the relevant specific material designation (taken from subrule .5B in the chapter...) e.g., wall charts, filmstrips versus microfiches (p. 288).

Cartographic materials include: atlas, diagram, globe, map, map section, profile, relief model, remote-sensing image, view (p. 108).

Manuscripts: not referred to as such except in other physical details (300\$b); same as book format, plus: items, ft. (ca. items OR v. OR boxes) (pp. 130-31).

Music materials: score, condensed score, close score, miniature score, piano [violin, etc.] conductor part, vocal score, piano score, chorus score, part (p. 150).

Sound recordings: sound cartridge, sound cassette, sound disc, sound tape reel, sound track film (p. 170).

Motion pictures and videorecordings: film cartridge, film cassette, film loop, film reel, videocartridge, videocassette, videodisc, videoreel (p. 190).

Graphic materials: art original, art print, art reproduction, chart, filmstrip, filmstrip, flash card, flip chart, photograph, picture, postcard, poster, radiograph, slide, stereograph, study print, technical drawing, transparency, wall chart (p. 209).

Computer files: computer cartridge, computer cassette, computer disk, computer reel [add new one SMDs as new physical carriers are developed] (p. 231).

Three-dimensional artefacts and realia: art original, art reproduction, braille, cassette, diorama, exhibit, game, microscope slide, mock-up, model.

If none of these terms is appropriate, give the specific name of the item...as concisely as possible—e.g., “hand puppet; jigsaw puzzle” (pp. 250-51).

Microforms: aperture card, microfiche, microfilm, microopaque.

Add cartridge, cassette, or reel, as appropriate...add cassette, if appropriate, to microfiche (p. 266).

MARC Formats: Book, Serial, AMC, MRDF, Maps, Music, Visual Materials. Separate out carriers of information—e.g., microform, digital—from rest.

The heart of the problem seems to be that any item falls into several different classifications simultaneously and orthogonally—a good time to remind ourselves that classification is just a way for human beings to think logically about the universe, which is a continuum. Following—in no particular order—are the major types of classification that a cataloger may have in mind:

a. manuscript	printed
b. monograph	serial
c. analog	digital
d. text	graphics
e. moving	still
f. codex	single-sheet
g. eye-readable	not eye-readable
h. sound	no sound
i. ascii	binary
j. actual object (3-D)	representation of an object (generally 2-D)

Lack of consistency—brought about, one suspects, mainly because the codex has predominated as a method of transporting information for much of the history of libraries—becomes obvious when one cruises through AACR2R seeking out 300\$a.

For nonatlas, nondigital monographs and serials: pp. or v. [carrier; general assumption that item is in one volume unless otherwise stated].

For cartographic materials: 1 map/view/section/diagram [“on 1 sheet” is assumed], 1 globe [“on 1 sphere” is assumed], 1 model (ambiguous—what kind of model?), 1 atlas (pagination), 1 diagram [on 1 sheet] (ambiguous—what kind of diagram?), 1 view [on 1 sheet], 1 section.

For digital data: we are firmly in carrier-land—“9.5.B1. When new physical carriers” appear, formulate terms—computer cartridge, computer cassette, computer disk, computer reel, computer laser optical disc (why is it that everywhere except in catalog records are these called CD-ROMs? Do most users even know what “computer laser optical disc” means?).

Going back to AACR2 for the sake of comparison, we are dealing with what seem more like intellectual forms: i.e., data file, program file, object programs. For contrast, the map curator of the Royal Library, National Library of the Netherlands, includes in physical descriptions the number of floppies, files, and bytes as appropriate (J. Smits, personal communication, 16 December 1993).

This whole problem is especially noticeable when one, for example, has a serial CD-ROM of AVHRR data (Advanced Very High Resolution Radiometer)—it qualifies simultaneously as serial, machine-readable data

file, graphic (and so on), and map (since the data are generally displayed on a computer screen as a map); so, which SMD (or collection thereof) are you going to use? What takes precedence? Information type or carrier type? And why not be consistent throughout AACR (e.g., “____ intellectual item/substantive form of material on ____ physical form”) or isn’t it possible?

This brings us to the subject of file characteristics. This field is in the same family as 255, Mathematical data (scale; projection; coordinates).¹ In AACR2R (it did not exist in AACR2), file characteristics are given as:

Computer data (____ files, ____ records, ____ bytes)

Computer program (as above)

Computer data and program (as above).

It would seem more appropriate for this field to include, in addition to whether the material is data or software, such matters as whether it is text or graphics or both or whether data can be displayed in color (more about this later). Bytes are analogous to words or possibly to number of pages and therefore seem more appropriately placed in physical description (specifically in 300\$a); also, geospatial data sets are often large (e.g., a black and white 9" x 9" air photo scanned at 600dpi requires 29 megabytes, a color photo requires 91 megabytes, all bands of a SPOT image could require approximately 400 megabytes), and it makes sense to give these in megabytes rather than in bytes. Files are like chapters. What is the point of having this information? Most importantly, users have little or no interest in the number of chapters in a book, and I have about as much interest in counting files as I do in counting chapters or plates. And what does “record” mean? It is not in the glossary of AACR2R. As indicated, these seem more appropriately located in physical description than in file characteristics.

Reverting to physical description, for 300\$a, how about an amalgam of AACR2 and AACR2R: computer data on CD-ROMs or 580 megabytes on 1 computer laser optical disc or 10 images (580 megabytes) on 1 CD-ROM? But this will not work for GIS databases, where the number of maps that might be constructed is infinite. As always, everything depends upon the author’s or publisher’s intent.

Moving on to other physical details, 300\$b, we have: 1 computer laser optical disc : col. Although the intent is to indicate that the data on the disc will display in color, what is actually being said is that the disc itself is colored. Even the description: 580 megabytes on 1 computer laser optical disc : col. is not much better because of the way whether or not data may be displayed in color is carried:

10 images (580 megabytes) : col. is accurate. Bytes are not in color in the way that a map is in color. In raster data, each pixel is associated

with a value from 0 to 255, which denotes what grey-scale value that pixel has. The image may be displayed either as grey-scale (black and white) or as color depending on the software in use.

Yet another point. In the version of guidelines for cataloging Internet resources that I have, it is stated that one does not use 300 because there is not a physical item to describe. It would be more accurate to say that one does not have a physical carrier to describe. If one considers number of bytes (which definitely do take up space as anyone who has watched the memory in a pc go to 2 percent empty knows) to be analogous to pages, then 300\$a is an appropriate place for size of file.

PRODUCTION VERSUS PUBLICATION

If publication is defined as distribution of multiple copies by sale or by other transfer, then digital data seems to qualify. In the 1993 *Guidelines for Description of Internet Resources* (Patton, 1993), it is noted that, generally speaking, only electronic serials are considered published, but if a monographic item carries a formal statement similar to that found on the title page of a monograph, then it may be considered published (p. 1). Practically speaking, any file that is made available for others to use—whether through anonymous FTP site or Mosaic—is, in digital terms, published.

University faculty members are presently dealing with this “is-it-published” conundrum in the sort of way that wonderfully fixes one’s attention—is work made available over the Internet considered to be a publication that will “count” toward getting tenure?

MERGING TWO STANDARDS

Metadata in Project Alexandria must be compatible with two metadata standards, USMARC and the Content Standards for Digital Geospatial Metadata. USMARC is a database format; it is not a cataloging-rules standard—that standard is AACR. One step yet further away are the cataloging concepts upon which AACR is based. This may explain some of the problems I have seen in applying Content Standard for Digital Geospatial Meta Data, since making equivalencies between something that is a concepts standard (Content Standards for Digital Geospatial Metadata) and something that is a database-format standard (MARC) is obviously going to run into difficulties. At a more abstract level, the two standards are different. Both are concerned with accurately and briefly describing the item in hand, but the Content Standards for Digital Geospatial Metadata is intended mainly for use by data producers, while USMARC is mainly for agencies that make information available but generally do not produce it.

Another point is that USMARC has evolved since the late 1960s from cataloging policies and practice that date back 100 years and more, while the Content Standards for Digital Geospatial Metadata was put together in just a few years under the gun of a Presidential Executive Order stating that all federal agencies had to document their geospatial data using a standard that the Content Standards for Digital Geospatial Metadata would produce. These different beginnings engender different attitudes and result in documents that are different.

The much longer time frame in which current cataloging practice in libraries has had a chance to evolve has the additional benefit of allowing that practice to have a sturdy base of supporting standards, e.g., the USMARC: database format; ISBD; LCSH, and by extension; BGN place names; USGS lexicon of stratigraphic names; LC's schedules for classification; LC NAF; Z39.50; specifically for cartographic materials, AACCCM; and so on.

MULTILEVEL DESCRIPTION/PARENT-CHILD RELATIONSHIP

AACR proposes but LC disposes. In AACR2, multilevel description became an option for catalogers, including those who recognized users' needs for planetospatial data at the sheet or frame level—a problem, given that standard cataloging is at the series or flight level. For example, it is rare that people need to see every sheet of Morocco at 1:100,000, yet that is the only level at which the series appears in standard cataloging. It was only when USMARC finally had a linking field—772—that libraries were able to put this into practice. There is no note relating a child to a parent—the link is solely in the vertical relationship tag and is carried as a number. Mainly, it seems to have been applied in the AMC format although the then National Map Collection of Canada has used it (Parker, 1990). The Canadians also used two local-interest fields which unfortunately have no USMARC equivalents:

UTLAS field 1083

Local-interest code:

001 parent record

002 subrecord

003 both parent and subrecord

UTLAS field U035

UTLAS local information code: 1 parent record

2 analytic \$a accession no. (Parker, 1990, p. 89)

Possibly 008/25 (Cartographic material type), which is confined to: single map, map series, map serial, globe might be expanded to include

the categories given as 002 and 003 in UTLAS field 1083, since map series is, in effect, 001 (although such a series could also be a subrecord). The general rule is that what is common to all children is recorded at the first level. Each successive level contains only the information pertinent to that level and does not repeat what is at a preceding level. There are several different layers, all of which, mercifully, are not needed in all cases for various forms of planetospatial data:

- a. air-photo flight:

Flight	Flight	Line	Roll	Frame
--------	--------	------	------	-------
- b. satellite imagery:

Overall mission name	Satellite number	Scene ID
----------------------	------------------	----------
- c. map series:

Parent	Another parent	Subseries	Child
--------	----------------	-----------	-------
- d. GIS:
 - i. Parent Tile (geographic areas adjoining each other; analogous to topographic map sheets in a series)
 - ii. Parent Various themes/layers/coverages

There is rarely much common metadata among coverages since each generally has different lineage. It would also be difficult to devise a concept that survives from one GIS software to another. It will be interesting to see how this system works when one applies it to digital data and specifically to vector data.

It can happen that a child can belong to more than one series—i.e., it can have more than one parent. This is the sort of occurrence that causes database-software engineers to go grey before their time. In addition, it will work best from a database point of view if child records have different fields than do parent records, but that is not possible. For example, both parent and child records will have title fields, call number, and a few other fields in common. This was indicated by a list generated for Alexandria of the child fields for each of four major formats the project was working with—map; air photo; satellite image; and digital data set. Another problem that occurs is, if one is using 772\$, why bother anymore with 4xx/8xx? The tentative answer seems to be that one still may use 4xx/8xx for items in monographic series such as the U.S. Geological Survey folded-map series (GQ, I, MF, etc.). Still, it could come in handy for such subseries as the many Atlas of Mars subseries in the I series.

The best way to handle this seems to be the following:

- a. use a tree analogy, with root-branch-leaf relationship—root as parent, branch as subseries, leaf as child;
- b. have a field, as the Canadians did, that indicates what any given record is in relationship to other records;

c. link the fields with 772\$w; for example:

Level 1 (root)	Parent record	
	RSN92850219 [control number; 001]	
Level 2	Sub-group	Item
	RSN92850226	RSN92850223
	772\$w92850219	772\$w92850219
Level 3	Item	Item
	RSN92850228	RSN92850227
	772\$w92850226	772\$w92850226 (Parker, 1990, p. 88)

d. when a user calls up a child, have software call the parent record. When the user calls up a parent and then requests all the children, have the standard response "This retrieves more than 5,000 records and will take 20 minutes; would you like to refine your search?" if there are a large number of children/leaves.

Another matter of interest here is linking related materials—e.g., DEMs and DLGs to the USGS quadrangles from which the data were digitized—but this can be handled within existing cataloging fields. *Guidelines for Bibliographic Description of Reproductions* (Association for Library Collections & Technical Services, 1995) neatly sidesteps this matter and does not include digital forms of an item (e.g., a raster scan of a specific nautical chart is not a reproduction of the specific chart).

What to Catalog

For its prototype, Alexandria scanned any nondigital items that were cataloged thus creating a digital file. So what does one do? Create a catalog record for the digital file? Alexandria stopped with creating the digital file and did not catalog it, but there is no reason why one has to stop there and many reasons to continue on. One needs to create a record for the scanned object, but it should be closely connected to the record for the original item, perhaps as a version of it. As Barbara Tillett (1995) put it so well in her paper on multiple versions as a digital equivalent of the supremely useful "dashed-on" entry from pre-USMARC days.

Subject Headings

There are a couple of subjects that need some work. Satellites should be referred to in a consistent fashion. Currently, we have both: Landsat satellites and SPOT (artificial satellite). The latter looks the more logical of the two.

A primary part of any geospatial-digital-data reference question is, "Is the data raster or vector?" With the new Content Standards for Digital Geospatial Metadata fields, USMARC does have a field specifically to note

this, but given the importance of the question, perhaps we should also have a subject heading for each. “Linear topographical spaces” seems to be how vector data sets are presently referred to in LCSH. There does not seem to be a subject heading for raster data.

This is not a subject heading relating solely to digital geospatial data. It is included here on the grounds that so much imagery of other planets in the solar system is in digital form, and it was that data that brought this problem—which previously had irritated me but not enough to do anything about it—most forcibly to my attention. We do not deal consistently with planetospatial data of Earth in LCSH. For other planets we have: Venus (Planet), Jupiter (Planet), and so on. We do not have, and very much need, Earth (Planet). This means that when one catalogs a geologic atlas of the Earth, the only LCSH subject heading one may use is: Geology \$x Maps which plops it in with such works as the making and use of geologic maps.

Bounding Coordinates

With very few exceptions, every reference question for planetospatial data starts out with location. So, having quickly determined bounding coordinates are essential. For spatial data in digital form, this information is often in the header or derivable from inside the digital data. For aerial photographs, deriving it is a nightmare. We need software that determines these coordinates, and GIS and image processing software may have the answers needed.

Also, what about coordinates for other planets? USMARC and Content Standards for Digital Geospatial Metadata both assume Planet Earth coordinates. We need a field—perhaps an indicator—that signals what planet is meant.

Time

When time becomes a matter of metadata interest, then we must specify what kind of time is involved (local time? Greenwich time?), and we need a field that tells the reader of the metadata what kind of time is meant.

CONCLUSION: TIME OF CHANGE

Some years ago, the cataloging community was the most conservative part of the library world, the least willing to consider change. Then, in short order, we got MARC, shared online cataloging, and AACR2, all of which quickly eliminated anyone who was resistant to change (I can still remember catalogers who took early retirement rather than to deal with AACR2). The cataloging community has shown itself, especially over

the last five years, to be very responsive to change, willing to put together a set of rules for cataloging new formats (e.g., data available over the Internet) in short order and send these out to the wider community to be tried out. Certainly MARC has its faults, and certainly the library world is looking at next steps (e.g., SGML). However, it has been both extraordinarily useful and extraordinarily successful, being adopted in one form or another internationally, and there are perhaps 60 million USMARC records in existence. It is important to remember that MARC was formulated at a time when computer power was far less powerful and far more expensive than it is now, from whence came many of the coded fields (contained generally in 001-049). What would be most helpful now would be if the cataloging community would take on the very difficult task of reworking the cataloging rules so that they are even more based on cataloging information and not on cataloging the book form.

Speaking more specifically, over the last six months it has become increasingly obvious that, for a cataloger to catalog digital planetospatial data accurately and quickly, the cataloger must know a considerable amount about such data—how to load it, how to read headers, how to scan materials and do image processing on the resulting files, and so on. Alexandria depended on two very capable, hard-working geography graduate students, one specializing in raster data and one in vector. Even as a short-term situation, this did not work perfectly. One needs the combination of knowledge of cataloging policy and practice and knowledge of how to deal with the digital side in order to produce good catalog records.

NOTE

- ¹ Scale is not logically applied to digital spatial data sets; scale of hard-copy data used for inputting is not the same as the scale of a specific hard-copy map being cataloged. More tellingly, the implication here—that a digital planetospatial data set may be displayed at only one scale—is not correct. The equivalent information for digital data is resolution which, no matter at what scale the data are displayed on a screen, remains the same. For example, the horizontal spatial accuracy of 1:24K DEMs is 30 meters; the 1:250K DEMs have a spatial accuracy that is tied to latitude—3 arc-seconds up to 50 degrees North, 6 arc-seconds from 50 to 70 degrees North, and 9 arc-seconds above 70 degrees North.

REFERENCES

- American Library Association. (1982). *Anglo-American cataloguing rules, revisions* (2d ed). Chicago, IL: ALA.
- Association for Library Collections & Technical Services. (1995). *Guidelines for bibliographic description of reproductions*. Chicago, IL: American Library Association.
- Bretherton, F. P., & Singley, P. T. (1994). *Metadata: A user's view*. [S.l.]: IEEE. (0-8186-6610-2/94 c1994 IEEE)
- Canadian General Standards Board. (1994). *Geomatic data sets cataloguing rules*. Ottawa, Ontario: Canadian General Standards Board and Canadian Library Association.
- Cartmell, V.; Parker, V.; & Stibbe, H. L. P. (Eds.). (1982). *Cartographic materials: A manual of interpretation for AACR2*. Chicago, IL: American Library Association.

- Dodd, S. A. (1982). *Cataloging machine-readable data files: An interpretive manual*. Chicago, IL: American Library Association.
- Farrell, B. (1994). Digital cartographic information in the map library. *Association of Canadian Map Libraries and Archives Bulletin*, 91, 1-9.
- Frank, S. (1993). *Cataloging digital spatial data in the information infrastructure: A literature and technology review*. Orono, ME: U of Maine Dept. of Surveying Engineering/NCGLA.
- Gorman, M., & Winkler, P. W. (1988). *Anglo-American cataloguing rules* (2d rev. ed.). Chicago, IL: American Library Association.
- Graham, C. (1990). Definition and scope of multiple versions. *Cataloging & Classification Quarterly*, 11(2), 5-32.
- Guidelines for the arrangement and description of cartographic archives*. (1993). [Albany NY?]: Alabama Dept. of Archives and History, New York State Archives and Records Administration, New York State Library, Manuscripts and Special Collections.
- Hoogscarpel, A. (1994). *Guidelines for cataloging monographic electronic texts at the Center for Electronic Texts in the Humanities* (CETH technical report no. 1). New Brunswick, NJ: Center for Electronic Texts in the Humanities. Available from: <<http://cethmac.princeton.edu/Docs/catguid.ps>>
- Olson, N. B. (1992). *Cataloging computer files* (Minnesota AACR2 Training Series, No. 2). Lake Crystal, MN: Soldier Creek Press.
- Parker, V. (1990). Multilevel cataloging/description for cartographic materials. *Western Association of Map Libraries Information Bulletin*, 21(2), 86-96.
- Patton, G. (1993). *Guidelines for description of Internet resources*. Unpublished handout at the American Library Association Committee on Cataloging: Description and Access meeting.
- Proposal no. 93-4: Changes to the USMARC bibliographic format (computer files) to accommodate online information resources*. (1992). Available from gopher://marvel.loc.gov:70/00/.listarch/usmarc/93-4.doc
- Tillett, B. (1995). *Versions*. Prepared for the Metadata Workshop, Dublin, Ohio, March 1-3, 1995. Washington, DC: Library of Congress.
- Urbanski, V., with Chang, B. C., & Karon, B. L. (1992). *Cataloging unpublished nonprint material*. Lake Crystal, MN: Soldier Creek Press.

MICHAEL DOMARATZ

Finding and Accessing Spatial Data in the National Spatial Data Infrastructure*

Information about where a thing is or where an event takes place is an important factor in decisionmaking in both the public and private sectors. Spatial data provide a unique context for integrating disparate observations and evaluating competing options. Factors of location, distance, pathways, and other spatial relationships often must be considered when making decisions about economic ventures, environmental and health concerns, responses to emergencies, and other issues.

Public and private sector organizations have quickly realized the usefulness of spatial data in their activities. The nation spends billions of dollars annually on just the collection, management, and dissemination of spatial data. Difficulties in finding and accessing data, and a lack of data documentation, hinder the spatial data community's efforts to work together and leverage this large investment. Through the National Spatial Data Infrastructure, government agencies, private companies, and nonprofit organizations cooperate to develop consistent, reliable means to share spatial data.

THE NATIONAL SPATIAL DATA INFRASTRUCTURE

The National Spatial Data Infrastructure provides a base or structure of relationships among data producers and users that facilitates data sharing. More formally, it is "the technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data" (Executive Office of the President, 1994).

The characteristics of the spatial data community greatly influence the approach to developing a "national" infrastructure. The many organizations in the community—including local, regional, State, and Federal government agencies, private companies, and non-profit organizations, have different (and sometimes competing) purposes, abilities, policies, interests, and needs. The many scientific or occupational disciplines in the community have different organizing principles, values, techniques, and terminologies. Some disciplines have a long experience with spatial data. Use of spatial data is quite new in others. All have something to

*This article is exempt from United States Copyright.

contribute to the community, although that fact is not always readily apparent to all members of the community! Finally, the members of the community are dispersed geographically, an important factor in evaluating means to link data users and producers. These and other characteristics of the community play a critical role in developing strategies to move forward.

"Spatial data" (sometimes called "geospatial data") identify "the geographic location and characteristics of natural or constructed features and boundaries on the earth" (Executive Office of the President, 1994). Most people would readily identify digital and paper maps, aerial photographs, and remotely sensed images as sources of spatial data. There also are many other types of data, including socioeconomic and demographic statistics, surveys of natural resources, and photographs and videotapes of landscape, that describe the locations and characteristics of geographic features. The spatial component of these data, which might be encoded using geographic coordinates such as longitude and latitude, a street address, or a county name or code, provides a key by which different information sources can be integrated and processed. The infrastructure must accommodate these different data so that potential users can find, evaluate, and integrate them.

An important aspect of the "infrastructure" is technology. Advances in computerized approaches to collect and process spatial data, and decreasing costs for using this technology, have helped spread the use of digital spatial data. Technologies such as the Global Positioning System, geographic information systems, and image processing help organizations that now use spatial data to do so more efficiently and effectively and entice other organizations to use these data for the first time. New, dynamic forms of spatial data are being created. Integration and use of data may result in new data being created.

Advances in telecommunications such as the Internet provide the ability to disseminate these digital data to a large audience. Before this technology was available, many organizations that collected and used spatial data did not have the printing, warehousing, and shipping infrastructure needed to distribute spatial data. The Internet now permits these organizations to make their information widely available, as well as to locate needed data that are produced by others. Traditional relationships within the community are changing rapidly as technology enables the emergence of new data producers and users, and new opportunities for collaborative data collection and use.

Work on infrastructure requires attention to other links within the community. Concerns and views vary widely on issues such as recovering the costs of data collection, freedom of information, and liability. Development of the infrastructure also depends on ensuring that new profes-

sionals are trained well, and that existing professionals keep up with the rapid technological change.

THE NATIONAL GEOSPATIAL DATA CLEARINGHOUSE

One challenge brought on by the changes in the community is that of finding and accessing digital spatial data. The amount of data being produced, the number of organizations producing data, and the decentralization of data production and distribution are growing. Distinctions between data producers and users are being blurred. These data often are not "published" in a traditional way. Mechanisms for finding and accessing information must accommodate these changes in the community.

These concerns are not unique to the spatial data community. Marchionini and Maurer note that "one clear difference between traditional libraries and digital libraries is that digital libraries offer greater opportunity for users to deposit as well as use information" (Marchionini and Maurer, 1995, p. 73). Wilensky provides another view: "For digital libraries to succeed, we must abandon the traditional notion of 'library' altogether. The reason is as follows: The digital 'library' will be a collection of distributed information services; producers of material will make it available, and consumers will find it and use it, perhaps through the help of automated agents" (Wilensky, 1995, p. 60).

Working with other members of the community, the Federal Geographic Data Committee is encouraging the development of the National Geospatial Data Clearinghouse as a means for the community to find and access digital spatial data. The clearinghouse is a referral service to discover who has what data. Designed with the decentralized distribution of data producers and users in mind, the clearinghouse is comprised of a set of information stores that use computer hardware, software, and telecommunications to link producers and users. To participate in the clearinghouse (see figure 1), producers create descriptions (or "metadata") of their data and make these descriptions available through the Internet. The resulting form of the clearinghouse is the "constellation" of sites, linked through the Internet (see figure 2).

Producers also may make their spatial data available directly through the clearinghouse. Many government organizations are taking advantage of this option to disseminate data that are in the public domain. Use of this option by other organizations will grow as methods for commercial transactions on the Internet mature.

To find and access data, a user communicates with the sites through the Internet, retrieves the metadata, and evaluates the metadata to determine the usefulness of available data for the planned application. When

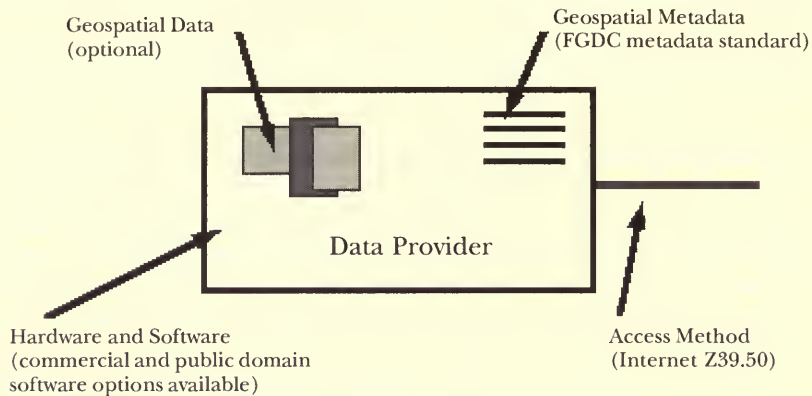


Figure 1. Components of a clearinghouse site.

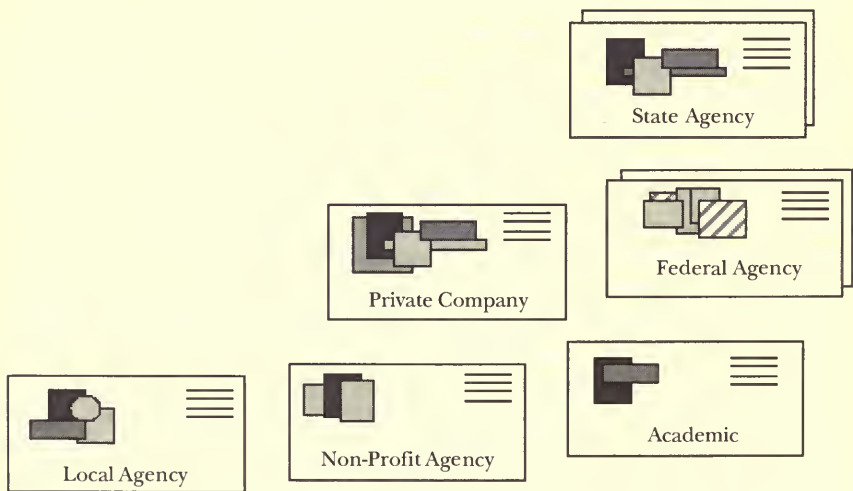
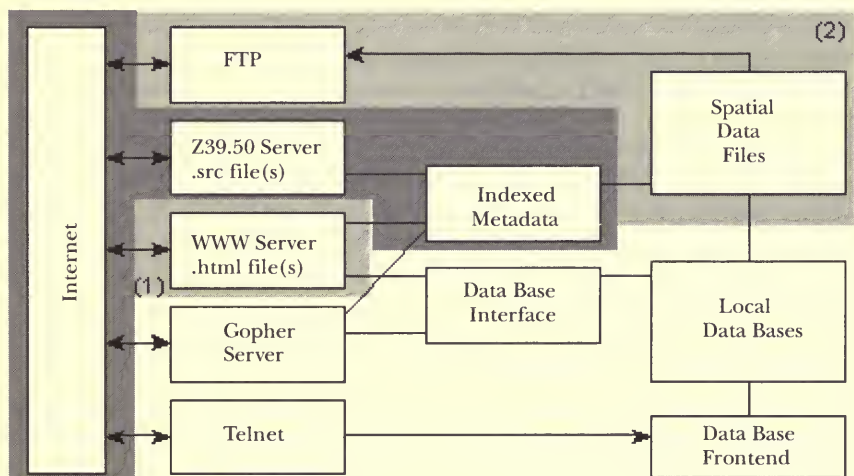


Figure 2. The clearinghouse is a “constellation” of sites from which data producers provide information.

useful data are identified, the user follows the instructions in the metadata for retrieving or ordering the data.

Another perspective on the clearinghouse is to consider different ways through which spatial data can be accessed using the Internet (see Figure 3). Of these alternatives, the clearinghouse is based on data producers providing access to metadata through Z39.50-compliant servers. Popular options include providing access to metadata through World Wide Web servers, and providing access to spatial data through File Transfer Protocol (FTP) services.

This approach to building the clearinghouse has been well received in the spatial data community. Producers value the ability to provide information about their data directly to potential consumers. The approach also encourages producers to keep the metadata current. Users appreciate the ability to access metadata and spatial data from their desks, and the ability to determine for themselves which data are the most suitable for their applications. Federal agencies, required to participate in



Basic configuration of the clearinghouse.

Option World Wide Web access to metadata (1) and File Transfer Protocol access to spatial data (2)

Figure 3. Of the methods through which the Internet can be used for spatial data, the clearinghouse uses Z39.50-compliant servers to provide metadata. Popular options include the use of the World Wide Web and the File Transfer Protocol.

the clearinghouse by Executive Order 12906, have been pleased with the initial response to their efforts. For example, the U.S. Fish and Wildlife Service reported that, in the first month of operation, approximately 29,000 digital maps from the National Wetlands Inventory were retrieved. The U.S. Geological Survey reported a similar volume of interest, with 40,000 files of spatial data retrieved in the first three months of operation (Federal Geographic Data Committee, 1994a).

The implementation of the clearinghouse is just beginning, and there are many challenges ahead. It has been noted that "the Internet is starting to provide the largest library humankind has ever had. As true as this may be, the Internet is also the messiest library that ever has existed" (Marchionini & Maurer, 1995, p. 72). This concern has been noted about the clearinghouse, and the Federal Geographic Data Committee is working within the spatial data and Internet communities to develop means to find and evaluate data more efficiently. New tools for using the Internet will result in new ways to implement the clearinghouse. The committee also sponsors a competitively awarded cooperative agreement program to encourage collaborative experimentation and implementation of the clearinghouse.

CONTENT STANDARDS FOR DIGITAL GEOSPATIAL METADATA

The lack of documentation for existing spatial data also hinders the spatial data community's ability to leverage its data investments. Many times organizations find data that seem to be useful for an application only to discover that very little is known about the data. For most organizations, concern about failure and liability is too great to risk the use of data that are not documented.

Data documentation, or metadata, describe the content, quality, condition, and other characteristics of data. Metadata for spatial data include:

- *Identification Information*—basic information about the data. Examples include the title or other identifier, the geographic area covered, currentness, and rules for acquiring or using the data.
- *Data Quality Information*—an assessment of the quality of the data. Examples include positional and attribute accuracy, completeness, logical consistency, and lineage (the sources of information and methods used to produce the data).
- *Spatial Data Organization Information*—identification of the mechanisms used to represent spatial information in the data. Examples include the method used to represent spatial positions directly (such as raster or vector) and indirectly (such as street addresses or county codes), and the number of spatial objects in the data set.

- *Spatial Reference Information*—description of the reference frame for, and means of encoding, coordinates in the data. Examples include the name of, and parameters for, map projections or grid coordinate systems, horizontal and vertical datums, and the resolution of the coordinate system.
- *Entity and Attribute Information*—information about the thematic content of the data, including the entities types, their attributes, and the domains from which attribute values may be assigned. Examples include the names and definitions of features, attributes, and attribute values.
- *Distribution Information*—information describing how to obtain the data. Examples include the means available to contact a distributor, available data formats, information about how to obtain data online or on physical media (such as cartridge tape or CD-ROM), and fees for the data.
- *Metadata Reference Information*—information about the metadata. Examples include the date the metadata were created and the means to contact the organization that created the metadata.

The data producer is the best source of this information. Details about the boundary of the area encoded in the data, the quality of the data, coordinate systems, data dictionaries, and other elements of metadata are all available when spatial data are produced. The best time to collect metadata is when the data are being collected.

Can data producers be persuaded to collect metadata? Fortunately, a major use of metadata is of great interest to data producers. Metadata provide a means to organize and maintain a producer's internal investment in data. Metadata help organizations to insure themselves from loss of knowledge about their data caused by personnel changes or by the passage of time. Metadata also help to protect organizations from conflicts caused by misuse of data.

In addition to maintaining internal investments in data, there are two other uses of metadata important to the spatial data community: (1) enabling participation in data clearinghouses, and (2) supporting transfers of data. Metadata are the core information of a data clearinghouse. Through the clearinghouse, an organization can find useful data that are available from others and make its data known to new customers. An organization also can identify other organizations with similar interests that may be potential partners in data collection and maintenance activities. Metadata also are essential information during the transfer of spatial data between organizations. For data to be useful to an organization, the organization must be able to integrate them into its holdings and applications. Metadata provide critical information needed to process and ingest new data.

As the spatial data community recognized the value of metadata, interest in standards for metadata grew. The Federal Geographic Data Committee sponsored a forum on the subject in 1992. At the forum, the participants heard descriptions of different approaches to setting standards for metadata, and agreed on the need for a standard on information content for metadata for digital spatial data. Volunteers drafted a standard which the Federal Geographic Data Committee offered for public review from October 1992 to April 1993. Extensive comments were received from the public. The committee revised the draft based on these comments and tests conducted by its member agencies. The committee also coordinated its efforts with those for related activities, such as the Machine Readable Cataloging (MARC) and the Government Information Locator Service (GILS). The standard was approved by the committee on June 8, 1994 (Federal Geographic Data Committee, 1994b). Geographic information coordination committees in several states also have adopted the standard.

The standard supports the common uses of metadata: to enable an organization to protect its internal investments in data, to support data clearinghouses, and to support data transfer. The standard provides for the encoding of information needed to satisfy common uses of metadata:

- availability—data needed to determine the sets of data that exist for a geographic location.
- fitness for use—data needed to determine if a set of data meets a specified need.
- access—data needed to acquire a set of data.
- transfer—data needed to process and use a set of data.

The standard provides a common set of terminology and definitions for the documentation of spatial data. The standard establishes names for data elements and groups of data elements, the definitions of these data elements and groups, and information about the values that are to be provided for the data elements. Information about elements that are mandatory, mandatory under certain conditions, and optional (provided at the discretion of the data producer) also is provided.

The standard specifies the information content of metadata, but does not specify how this information is organized in a computer system or in a data transfer, nor the means by which this information is transmitted or communicated to the user. The variety of means for organizing data in a computer or in a transfer, the different institutional and technical capabilities of data producers, and the rapid evolution of means to provide information through the Internet provided the basis for this decision.

Recognizing the different needs and abilities within the spatial data community, the standard provides leeway in a number of implementa-

tion decisions. Elements of metadata can be encoded for different levels of granularity of data, ranging in size from large collections of data files to individual lines and areas. The amount of detail that is encoded may vary. Different data may vary in their significance or value, and the effort expended in developing metadata should correspond to the value of the data. Decisions about documenting existing and new data should be considered carefully. Organizations with large holdings of undocumented "legacy" data are concerned about the costs of documenting these holdings, and sometimes allow the consequences of past practices to control their decisions about documenting new data.

The Federal Geographic Data Committee recognizes the need for the standard to evolve with the changing needs of the community, and is working with the community on improvements. Current activities include identifying a core set of metadata to facilitate searches, developing means of providing "lite" amounts of metadata, and developing means of adding locally used extensions to the standard.

FUTURE DIRECTIONS

The factors which have fashioned activities to develop the National Spatial Data Infrastructure will continue to challenge current views about how to collect and share digital spatial data. Successful approaches will be those that allow the community to contribute, share, integrate, and use spatial data for varying units of space, periods of time, and thematic detail. It is difficult to know exactly what may emerge from this dynamic environment. Chrisman (1994) has identified some things that a digital library of geographic information should *not* be: it is not a collection of map sheets; it is not just one snapshot in time; and it is not modeled on a digital library of books or scientific publications.

The Federal Geographic Data Committee sponsors projects to develop the National Spatial Data Infrastructure and to improve the community's ability to work together. For more information about the committee's activities, visit the committee's World Wide Web site at <URL: <http://fgdc.er.usgs.gov>>, or contact the committee by electronic mail at gdc@usgs.gov or by postal mail at the FGDC Secretariat, c/o U.S. Geological Survey, 590 National Center, Reston, Virginia 22092, USA.

REFERENCES

- Chrisman, N. (1994). A vision of digital libraries for geographic information, or how I stopped trying to find the on-ramps to the information superhighway. *Geo Info Systems*, 4(4), 21-24.
- Executive Office of the President. (1994). Coordinating geographic data acquisition and access: The National Spatial Data Infrastructure (Executive order 12906). *Federal Register*, 59(71), 17671-17674.

- Federal Geographic Data Committee. (1994a). *Content standards for digital geospatial metadata*. Reston, VA: Federal Geographic Data Committee.
- Federal Geographic Data Committee. (1994b). *The National Geospatial Data Clearinghouse: A Report on federal agency activities within the First Six Months of Executive Order 12906*. Washington, DC: Federal Geographic Data Committee.
- Marchionini, G., & Maurer, H. (1995). The roles of digital libraries in teaching and learning. *Communications of the ACM*, 38(4), 67-75.
- Wilensky, R. (1995). UC Berkeley's digital library project. *Communications of the ACM*, 38(4), 60.

WILLIAM E. MOEN

The Government Information Locator Service: Discovering, Identifying, and Accessing Spatial Data

The Government Information Locator Service (GILS) is a new federal initiative to assist the public in discovering, identifying, and locating government information. GILS may play a special role in helping the spatial data community to search and retrieve information about spatial information resources created, collected, or held by federal agencies. GILS is a decentralized standards-based approach to network discovery and retrieval. The basic components of GILS are structured records (i.e., the GILS Locator Records) with standardized data elements that describe and provide access information to federal information resources; agency-based information servers hosting the Locator Records; client software to initiate information retrieval transactions; and ANSI/NISO Z39.50 as the communications protocol between clients and servers. This paper provides an overview of GILS and discusses ANSI/NISO Z39.50, the American National Standard for information retrieval, and its use in GILS. The paper concludes with a discussion of the implications of GILS for the discovery and use of spatial data.

INTRODUCTION

A major barrier in the effective use of distributed electronic networks and the information available through them is the lack of adequate mechanisms for discovering and retrieving information that is relevant to users' information needs. Nowhere is this more true than in the realm of federal information resources and more specifically geographic and spatial data created, collected, or held by federal agencies.

The Government Information Locator Service (GILS) is one mechanism that will assist users to search for and retrieve information. GILS is a decentralized standards-based approach to network discovery and retrieval and uses ANSI/NISO Z39.50, the American National Standard for information retrieval. GILS is a network-accessible service through which the public will be able to identify and locate federal information resources. GILS will help users identify what resources exist and provide information about those resources to allow the users to determine the utility of the information resource for their needs. Additionally, GILS provides

users the necessary information for accessing or acquiring the resources. While the focus of the initial GILS effort has been on federal information resources, its use is not limited to those resources. Some state governments and other organizations are investigating the use of GILS to assist their user communities in discovering and retrieving information.

This paper provides an overview of GILS and discusses Z39.50 and its use in GILS to facilitate information retrieval in the increasingly distributed electronic environment of federal information. The paper concludes with a discussion of the implications of GILS for the discovery and use of spatial data.

BACKGROUND AND CONTEXT

GILS is a new federal initiative to assist users in discovering, identifying, and accessing or acquiring federal information resources. Through the use of existing technology, GILS provides a framework and specifications for agencies to describe their information resources and make those descriptions available to users directly via the Internet or through any number of intermediaries. Additionally, GILS is intended to assist government agencies in the management of federal information resources. Adequate management of information resources, enhanced access to public information, and dissemination of agency information products are three primary goals of GILS. It is important to see these goals as information policy goals and to understand that GILS is a policy-driven initiative.

A brief overview of the current policy and technology context within the federal government will assist in understanding the emergence of GILS. For many years, the primary policy instrument for federal information has been the Office of Management and Budget's (OMB) Circular A-130, "Management of Federal Information Resources." Originally published in 1985, A-130 directed federal agencies in their information management activities. Beginning in the late 1980s, OMB began the process of revising A-130. This effort culminated in the publication of substantial revisions to the circular in 1993 and in 1994 (Office of Management and Budget, 1994b). While maintaining a focus on federal information resources management, sections of the revised circular also addressed agencies' responsibilities in providing access to, and dissemination of, government information. In part, the revised A-130 set the stage for GILS.

Other activities have also shaped the policy context of the current GILS initiative. For example, two research studies examined problems and design issues related to a federal information locator system (McClure et al., 1990; McClure et al., 1992). Those studies suggested that a user-based approach to designing a locator system, policy analysis and policy advocacy, and an awareness of technology trends and information technology standards can be particularly effective to connect the needs of in-

formation users and providers in achieving broader federal information policy goals.

The development of GILS occurred within the larger policy context reflected, for example, in the National Information Infrastructure (NII) and the National Performance Review (NPR) initiatives. The NII and NPR are two examples of important policy statements that reflect the Clinton administration's encouragement of the government's use of electronic networks to increase public awareness of, and access to, government information and services. The NII initiative encourages agency efforts to expand electronic dissemination through diverse channels (Information Infrastructure Task Force, 1993). The GILS is seen as a federal contribution to the NII by providing a "virtual card catalog" to government information. The National Performance Review, released by Vice President Gore in September 1993, encourages agencies to develop electronic dissemination programs that employ the right balance of public and private sector efforts (National Performance Review, 1993). The current incarnation of OMB Circular A-130, however, provides important enabling policy for GILS. The circular directs agencies much more forcefully in their responsibilities in the areas of information access and dissemination. Agencies are, among other things, created to:

- help the public locate government information maintained by or for the agency;
- disseminate information on equitable terms;
- develop aids to locating information;
- establish and maintain inventories of all agency information dissemination products;
- use electronic media and formats to make government information more easily accessible and useful to the public; and
- use information technology standards to create open and interoperable information systems.

GILS is a policy and technology response to the needs of the public to identify, locate, and access or acquire government information. OMB recently formalized the policy on GILS in its Bulletin 95-01, "Establishment of Government Information Locator Service" (Office of Management and Budget, 1994a). The bulletin directs federal agencies to inventory their information resources and begin describing those resources in GILS Locator Records.

Before users can access or acquire information, they must determine whether that information exists, whether there are restrictions on its use, and what methods are available for accessing or acquiring it. A locator serves exactly that purpose. According to McClure, Ryan, and Moen (1992), a locator is a "machine-readable database that identifies different information resources (e.g., databases, libraries, clearinghouses, print

publications, bulletin boards, etc.) and describes the information available in these resources. Usually, the locator does not provide the actual information, but rather points the user to the information sources that do provide the needed information" (p. 2). Locator information is metadata or data about data. A locator is a point of entry for locating government information, regardless of the format and content of that information, and tells the user: (1) what information is available on a particular topic, (2) where that information is located, and (3) how the user would access that information.

AN OVERVIEW OF GILS

GILS is an agency-based, Internet-accessible, locator service. Direct users of GILS will connect to GILS servers via the Internet to find information about a wide range of federal information resources. The public will use GILS either directly or through intermediaries (the intermediaries obtain GILS information as direct users themselves or from other intermediaries [see Christian, 1994, p. 4 for a description of the various classes of GILS users]).

Federal agencies will develop and maintain GILS servers. These decentralized agency-based GILS servers enable ongoing maintenance responsibilities to be carried out by those who understand and manage information resources.

GILS servers are machine-readable databases that contain Locator Records describing federal information resources. A Locator Record consists of a number of data elements (i.e., GILS Core Elements) that identify and describe an information resource (GILS Core Elements are noted in uppercase letters throughout this paper). Several data elements are included in Locator Records to facilitate GILS navigation and network-based access to information. For example:

- Each Locator Record contains a **CONTROL IDENTIFIER** in the form of a Uniform Resource Identifier (URI). An agency's server may contain Locator Records with **CONTROL IDENTIFIERS** that identify Locator Records from other agencies' servers. This data element allows GILS Locator Records to be replicated on multiple servers for the convenience of GILS users.
- Each Locator Record contains an **AVAILABILITY** element that informs the user how to procure the described information resource. If the information resource is an electronic information system or electronic document, the **AVAILABILITY** element includes **AVAILABLE LINKAGE** information in both human- and machine-readable form. The network linkage information may be used to connect to, and access, the electronic information resource.

Different agencies may create or offer Locator Records describing the same information resource (these may be existing Locator Records that have been replicated and/or modified or entirely new Locator Records). These multiple records can offer different views of a single resource from the particular perspectives of the agencies creating/modifying a Locator Record.

The semantics of the Locator Records coupled with client software that understands these semantics and building upon the ability of the Z39.50 protocol to provide a uniform interface to multiple autonomously managed servers provide the user with the impression of seamless navigation among these distributed servers. The semantics of the Locator Records facilitate elimination of duplicate records further fostering the impression of a single system built out of autonomous distributed servers.

Each GILS server can be represented by a Locator Record in other GILS servers. Some of these servers will include references to all other GILS servers, and these might be regarded as a kind of "directory of directories." However, GILS itself does not assign any hierarchical status to specific servers nor does it specify a "root server." Rather, the structure and content of the GILS Locator Records enable, for example, the aggregation of Locator Records in "directories" that could be offered by one or more federal agencies or other organizations.

A GILS server accessed using Z39.50 in the Internet environment acts primarily as a pointer to information resources. GILS servers will support searching (i.e., accept a search query and return a result set or diagnostic messages) and may support browsing (i.e., accept a well-known search query and return a list of Locator Records in brief display format). Direct users must have prior knowledge of at least one GILS server and its network address and must be able to access it to enter the GILS. Once connected to a GILS server, users supported by appropriate clients that understand the GILS Profile, may navigate through single or multiple GILS servers by following the links provided in the Locator Records. The use of the national standard for network information retrieval, Z39.50, provides for interoperability between clients and multiple servers. GILS, then, is a distributed resource consisting of agency-based servers accessible via Z39.50 that provide users with the potential to discover, identify, and locate federal information resources.

THE GILS PROFILE

The Federal Information Processing Standards (FIPS) Publication 192 lays out the specifications for GILS applications (National Institute of Standards and Technology, 1994). A profile is "a set of one or more base standards, and where applicable, the identification of chosen classes, sub-

sets, options and parameters of those base standards, necessary for accomplishing a particular function" (International Organization for Standardization/ International Electrotechnical Commission, 1992, p. 2). Profiles are also referred to as "functional standards," "implementation agreements," or "specifications." Since open systems standards often include choices and options, profiles specify the values and parameters of a standard for an application to increase the likelihood of interoperability and interworking of separate implementations. A profile, then, is a set of implementation agreements that guide implementors in applying one or more standards in a specific and limited context. Separate implementations will have an improved likelihood of interoperability and interworking when they conform to a common application profile.

The development of the GILS Profile was completed in 1994 under the auspices of the research project, "Expanding Research and Development on the ANSI/NISO Z39.50 Search and Retrieval Standard," coordinated by Syracuse University and the United States Geological Survey and funded by the Interagency Working Group on Data Management for Global Change. A project team comprising experts in Z39.50 implementations, system implementations, and information organization, and representatives of federal agencies developed the specifications (i.e., the GILS Profile) for initial GILS implementations (for a complete description of the research project and the development of the GILS Profile, see Moen and McClure [1994]).

The GILS Profile development project built upon a previous study, *Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators* (McClure et al., 1992). That study, which was conducted for OMB, the National Archives and Records Administration (NARA), and the General Services Administration (GSA), recommended that each federal agency establish a network-accessible locator that describes its information resources. The study also recommended that agencies use Z39.50 as the appropriate information retrieval protocol to achieve a distributed standards-based Government Information Locator Service.

The *Government Information Locator Service* (Christian, 1994) provided the project team with high-level requirements for GILS. Based on those requirements, the project team delineated assumptions about the operation and information flows of GILS and developed functional requirements. This process allowed the project team to identify a subset of Z39.50 and other existing and emerging standards that would support these functional requirements.

The GILS Profile fully specifies the use of Z39.50 for the GILS application. The profile also addresses other aspects beyond Z39.50 for information servers that are GILS conformant. The current version of the

profile only addresses servers, and while acknowledging a client's role in information retrieval, the Profile does not specifically address client functionality. A GILS client, however, will be able to interoperate with any GILS server. While the GILS Profile specifies many aspects of GILS applications, it does not address information system characteristics such as interface requirements, the internal structure of databases that contain GILS Locator Records, or search engine functionality.

Z39.50 provides a key part of the foundation for GILS. This standard enables the interoperability of a variety of systems and hardware platforms in a client/server environment for the purposes of information retrieval.

Z39.50: A COMPUTER-TO-COMPUTER PROTOCOL FOR INFORMATION RETRIEVAL

Understanding how GILS works requires some familiarity with ANSI/NISO Z39.50-1992, *Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection* (National Information Standards Organizations, 1992). Michael and Hinnebusch (1995) argue that "Z39.50 is the single most important networking standard available today" (p. 15) and state that: "As far as the library community is concerned, it is the most important protocol available today" (p. 21). Z39.50 provides users with the capability to search and retrieve information in the networked environment. Z39.50 recognizes that information retrieval consists of two primary components—selection of information based upon some criteria and retrieval of that information. It provides a common language for both activities. Z39.50 standardizes the manner in which the client and the server communicate and interoperate even when there are differences between computer systems, search engines, and databases.

The National Information Standards Organization (NISO), an American National Standards Institute (ANSI) accredited standards developer that serves the library, information, and publishing communities, approved the original standard in 1988 (referred to as Z39.50-1988 or Version 1). Shortly after the approval of the standard in 1988, a group of Z39.50 implementors began work to enhance and expand the utility of the standard. NISO balloted a revised version of Z39.50 and published the new standard in 1992 (referred to as Z39.50-1992 or Version 2). Continuing development of the standard by the Z39.50 Implementors Group (ZIG) resulted in the third version of the standard (Version 3). A NISO ballot of Version 3 will be completed in the first quarter of 1995.

Oriented initially toward information retrieval of bibliographic records, Z39.50 is not limited to the formats and kinds of data it can handle. New features proposed in Version 3 extend Z39.50 to support full text and images and provide the functionality to accommodate information

retrieval beyond simple bibliographic information. These features emphasize the standard's utility in complex information infrastructures that process and handle information in various formats.

Z39.50 can be implemented on any platform and enables different computer systems—with different operating systems, hardware, search engines, database management systems—to interoperate and work together seamlessly, and thus supports an open systems environment. Z39.50 supports information retrieval in a distributed, client and server environment where a computer operating as a client submits a search request (i.e., a query) to another computer acting as an information server. Software on the server performs a search on one or more databases and creates a result set of records that meet the criteria of the search request. The server returns records from the result set to the client for processing. In a client/server architecture, software for end-user interaction and display (the client) is separate from the software that manages the information, performs the search, and returns the results (the server). Z39.50 does not address the user interface (e.g., its “look and feel”), but there are protocol specifications and procedures that pertain to the Z39.50 client (referred to in the standard as the Z39.50 “origin”) such as the initiation of an information retrieval query and how it requests specific operations from the server. There are also protocol specifications and procedures that pertain to the Z39.50 server (referred to in the standard as the Z39.50 “target”) such as the manipulation of result sets and the formats in which it returns records to the client. Z39.50 addresses the complex communication among computers for the purposes of information retrieval.

Each database residing on information servers can have unique characteristics. For example, databases may differ in the way they store data and in the access points available for searching. The records in each of the databases may also have different structures and consist of different data elements. The objective of Z39.50 is to support computer-to-computer communication in standard and mutually understandable terms and support the transfer of data between the systems independent of the structure, content, or format of the data in a particular system. However, in individual implementations, servers may be limited to specific formats of data that can be exported and the access points that are supported for searching.

When a database is searched, the client passes a query to the server. The query contains search “terms” (e.g., terms that the user has identified to be matched against access points in the database) and “attributes” of those search terms (e.g., specifying the terms as an “author” or “title,” specifying if the terms are to be “truncated,” etc.). Queries can include different attribute types. For example, if a user wants to search for an author's name, a “use” attribute specifies the search term as “author.” If the user wants to search for

all books published after a certain date, a “use” attribute specifies the search term is a “date of publication” and a “relation” attribute specifies that the user wants all dates of publication “greater than” a particular date. Z39.50 enumerates these attribute types and their values in registered attribute sets. Standardized and mutually recognized attribute sets allow implementors a common basis for intersystem communication.

After the server executes a search of a database, the server creates a result set consisting of those records that match the criteria of the query. Clients can request that servers return records from a result set, or they can issue additional searches that further qualify a result set or use result sets as arguments in subsequent searches.

When the user wants to display records listed in the result set, ANSI/NISO Z39.50 provides choices about which data elements (i.e., element sets) from the database record the user can request. It also gives choices about the format for transferring the record (i.e., a record syntax) from the server to the client. Z39.50 registers standardized Element Set Names and Record Syntaxes to support client/server communication for this aspect of information retrieval.

Z39.50 is an information retrieval protocol. The standard grew out of an early recognition by people in the library community that users should be able to search remote information systems for information just as they search their own and without the need to learn new search commands and techniques. A Z39.50 implementation enables one interface to search for, and retrieve, information from multiple systems and thus provides end-users with nearly transparent access to other systems.

GILS AND Z39.50

There are several basic components of GILS—the data (i.e., machine-readable records) stored in one or more databases; an information server hosting the records; client software used to initiate a query; and Z39.50 as the communications protocol between the client and the server. Z39.50 distinguishes between Z39.50 functions that are the responsibility of the client and Z39.50 functions that are the responsibility of the server; these may be logically associated with the information server and the user’s client. The Z39.50 functions pertinent to this discussion are those related to searching, retrieving, and presenting GILS Locator Records.

GILS Locator Records

GILS Locator Records consist of a number of data elements. The data elements are used to contain information that identify and describe and provide access information to federal information resources. Figure

Title (M, NR)	Contact Street Address
Control Identifier (M, NR)	Contact City
Abstract (M, NR)	Contact State
Purpose (M, NR)	Contact Zip Code
Originator (M, NR)	Contact Country
Access Constraints (M, NR)	Contact Network Address
General Access Constraints (M, NR)	Contact Telephone
Originator Disseminator	Contact Fax
Control (O, NR)	Record Source (M, NR)
Security Classification Control (O, NR)	Date of Last Modification (M, NR)
Use Constraints (M, NR)	Record Review Date (O, NR)
Availability (M, R)	Agency Program (*, NR)
Distributor (M, NR)	Sources of Data (*, NR)
Distributor Name	Schedule Number (*, NR)
Distributor Organization	Controlled Vocabulary (O, R)
Distributor Street Address	Index Terms—Controlled (O, NR)
Distributor City	Controlled Term
Distributor State	Thesaurus (O, NR)
Distributor Zip Code	Local SubjectIndex (O, NR)
Distributor Country	Local Subject Term
Distributor Network	Methodology (O, NR)
Address	Spatial Domain (O, NR)
Distributor Hours of Service	Bounding Coordinates (O, NR)
Distributor Telephone	West Bounding Coordinate
Distributor Fax	East Bounding Coordinate
Resource Description (O, NR)	North Bounding Coordinate
Order Process (M, NR)	South Bounding Coordinate
Order Information (M, NR)	Place (O, R)
Cost (O, NR)	Place Keyword
Cost Information (O, NR)	Place Keyword Thesaurus:
Technical Prerequisites (O, NR)	Time Period of Content (O, R)
Available Time Period (O, R)	Time Period—Structured
Time Period—Structured	Time Period—Textual
Time Period—Textual	Cross Reference (O, R)
Available Linkage (O, NR)	Cross Reference Title (O, NR)
Available Linkage Type (O, NR)	Cross Reference Linkage (O, R)
Point of Contact for Further	Cross Reference Type (O, R)
Information (M, NR)	Original Control Identifier (O, NR)
Contact Name	Supplemental Information (O, NR)
Contact Organization	

M = Mandatory O = Optional R = Repeatable NR = Not Repeatable

* This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system.

Figure 1. GILS locator record core elements

1 lists the GILS core elements (for the complete semantics of each element, see the GILS specifications in FIPS 192).

Lynch (1992) argues that information semantics in a distributed computing environment must be addressed if the promise of networked information is to be realized. The GILS Profile's use of a common vocabulary

for data elements and a common information structure of the records is an important step forward. Accuracy, completeness, currency, and consistency of data in the records, however, will be criteria by which the quality of the data can be evaluated. While the technology (i.e., Z39.50 clients and servers) may be able to process locator information, the end-users will be badly served if the data are lacking in quality.

The National Archives and Record Administration (1995) has developed *Guidelines for the Preparation of GILS Core Entries*. The OMB stated in its Bulletin 95-01 that NARA should publish guidance for federal agencies on the content of GILS Locator Records to assist those agencies in creating high quality and useful records.

GILS Locator Records are machine-readable records stored in a database on an information server. The use of Z39.50 focuses less on the database record itself than on a standardized representation of the record for the common understanding of the client and server. One aspect of this common understanding is the available access points for searching the record (i.e., the GILS Attribute Set); another aspect is the elements of the record or those parts of the record that the client can request of the server to return to the client for display to the user (i.e., Element Set Names).

Searching GILS: The GILS Attribute Set

Z39.50 logically separates searching for information from retrieving information. Searching is done by formulating a query and passing that query to an information server. Retrieval is done by requesting the server return one or more records, or element of a record, in one or more record syntaxes to the client.

GILS servers may support a variety of search strategies including those:

- to find known items (e.g., where the user knows the exact TITLE of an information resource described in a Locator Record);
- to find resources whose Locator Records contain certain words or phrases;
- to find resources by topic (e.g., using a controlled vocabulary); and
- to find resources whose Locator Records meet other criteria (e.g., specific spatial data coordinates).

Searching in Z39.50 uses the concept of Attributes. "Use" Attributes are access points in a database record that can be searched. The searchable elements of GILS Locator Records correspond to GILS Use Attributes. The GILS Attribute Set is a superset of the Z39.50 Bib-1 Attribute Set and consists of all Bib-1 Attributes and additional specific GILS Use Attributes. Although GILS servers are required to support a minimal set of Use

Attributes, any of the GILS Use Attributes listed in Figure 2 could be used as access points for searching the GILS Locator Record. The exact manner by which the user constructs the query is an interface issue and not specified by the GILS Profile, but users supported by appropriate clients that understand the GILS Profile should be able to specify searches with each of the required Attributes. The extent of available access points or supported Attributes, however, will be implementation specific.

To ensure a minimal level of interoperability among the clients and servers, the GILS Profile requires that servers support a limited number of GILS Attributes (Figure 3 lists these required attributes). If a GILS server receives a query with any combination of these Attributes, it should process the query and never return any of the following diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Relation Attribute," or "Unsupported Attribute Type." Table 1 displays the combinations of required GILS Attributes.

Retrieval in GILS: The GILS Schema, Element Set Names, and Record Syntaxes

As a GILS server completes a search, it produces a result set and makes that available to a client. The GILS server provides the client with the contents of selected records from the result set using the Z39.50 Present Service. The GILS server must respond to requests that records be presented in any of three Record Syntaxes mandated by the GILS Profile and one of the four Element Set Names specified by the GILS Profile. The exact manner in which records are presented to the user is an interface issue and not within the scope of the GILS Profile. There are three important aspects for retrieving GILS Locator Records: the GILS Schema, the Element Set Names, and Record Syntaxes.

A schema "represents a common understanding shared by the [client] and [server] of the information contained in the records of the database represented by the schema" (National Information Standards Organization, 1995, p. 132). The schema describes and/or defines an abstract record structure for a database record and a tagSet that uses tagTypes and tags to represent the elements in a database record. A schema can represent the hierarchical structure of database records such as the structure of GILS Locator Records.

Z39.50 defines two basic tagSets (tagSet-M and tagSet-G), and these contain elements commonly found in many database records. The GILS Schema uses tags from tagSet-M and tagSet-G, and it also defines a GILS tagSet for elements in the Locator Record that do not correspond to tags already defined in tagSet-M or tagSet-G. There are two general classes of elements in the GILS Schema: (1) Primitive—elements that cannot have locally defined subelements (2) Constructed—elements that can have one or more subelements, any of which may be well-defined or locally defined

Required GILS Use Attributes			
Use #	GILS Attribute Name		
12	Local Number (Local Control Number)		
29	Local Subject Index (Local Subject Term)		
1005	Author-name Corporate (Originator)		
1012	Date/Time Last Modified (Date of Last Modification)		
1016	Any		
1019	Record Source		
2001	Distributor Name		
2002	Index Terms—Controlled		
Available GILS Use Attributes			
Use #	GILS Attribute Name	Use#	GILS Attribute Name
4	Title	2029	Contact Country
1007	Identifier-Standard (Control Identifier)	2030	Contact Network Address
62	Abstract	2031	Contact Hours of Service
2003	Purpose	2032	Contact Telephone
2004	General Access Constraints	2033	Contact Fax
2005	Use Constraints	2034	Agency Program
2006	Distributor Organization	2035	Sources of Data
2007	Distributor Street Address	2036	Thesaurus
2008	Distributor City	2037	Methodology
2009	Distributor State	2038	West Bounding Coordinate
2010	Distributor Zip Code	2039	East Bounding Coordinate
2011	Distributor Country	2040	North Bounding Coordinate
2012	Distributor Network Address	2041	South Bounding Coordinate
2013	Distributor Hours of Service	2042	Place Keyword
2014	Distributor Telephone	2043	Place Keyword Thesaurus
2015	Distributor Fax	2044	Time Period—Structured
2016	Available Resource Description	2045	Time Period—Textual
2017	Order Information	2046	Cross Reference Title
2018	Available Technical Prerequisites	2047	Cross Reference Linkage
2019	Available Time Period—Structured	2048	Cross Reference Type
2020	Available Time Period—Textual	2049	Original Control Identifier
2021	Available Linkage	2050	Supplemental Information
2022	Available Linkage Type	2051	Record Review Date
2023	Contact Name	2052	Originator DisseminatorControl
2024	Contact Organization	2053	Security Classification Control
2025	Contact Street Address	2054	Cost
2026	Contact City	2055	Cost Information
2027	Contact State	2056	Schedule Number
2028	Contact Zip Code		

Figure 2. GILS use attributes

by the record creator; string tags (i.e., text labels) identify locally defined subelements. Figure 4 presents a selection of the GILS tagSet.

Table 2 is a partial list of the GILS abstract record structure. The GILS tagSet identifies tags from tagSet-M with tagType 1 and tags from tagSet-G with tagType 2. GILS tags are identified with tagType 4. For

The GILS Use Attribute is listed followed by the GILS Use Attribute Number and the corresponding GILS Core Element names.

Use Attributes:

- Local Number (12; Local Control Number)
- Author-name Corporate (1005; Originator)
- Date/Time Last Modified (1012; Date of Last Modification)
- Record Source (1019; Record Source)
- Distributor Name (2001; Distributor Name)
- Index Terms—Controlled (2002; Index Terms—Controlled)
- Local Subject Index (29; Local Subject Term)
- Any (1016)

Structure

- Word (2)
- URx (104)
- Date (5)
- Word List (6)

Relation

- Greater than (5)
- Equal (3)

Figure 3. Required GILS Attributes

example, LOCAL CONTROL NUMBER is defined in tagSet-M as (1,14); TITLE is defined in tagSet-G as (2,1). ACCESS CONSTRAINTS is a specific GILS Locator Record element and is represented in the GILS schema as (4, 53). ACCESS CONSTRAINTS is constructed of subelements including SECURITY CLASSIFICATION CONTROL, which is represented in the schema as (4,27). The schema shows the hierarchical structure of ACCESS CONSTRAINTS and SECURITY CLASSIFICATION CONTROL through the Tag Path of (4,53)/(4,27).

Schemas may be used with a particular record syntax, the Generic Record Syntax (GRS) (see below). Z39.50 implementations that use GRS-1 allow the client to request that the server return specific elements of the database record. Since the GILS Locator Record is represented by the abstract record structure in the GILS Schema, all elements are structured and identified for processing.

For example, a user has submitted a query looking for all Locator Records that describe information resources created by a particular agency (i.e., ORIGINATOR). The server creates a result set. The client can then ask to have the result set records returned, and it can specify that the server should return in GRS-1 only the following:

- Control Identifier (4,1)
- Originator (4,52)
- Place Keyword (4,71)/(4,92)/(4,13)
- Place Keyword Thesaurus (4,71)/(4,92)/(4,14)

Being able to specify parts of the record provides the user with additional control over the information retrieval transaction.

TABLE 1
RECOGNIZED AND SUPPORTED COMBINATIONS OF GILS ATTRIBUTES

<i>Use</i>	<i>Structure</i>			<i>Relation</i>	
	<i>Word</i>	<i>URx</i>	<i>Date</i>	<i>Word List</i>	<i>Greater Than</i> <i>Equal</i>
Local Number	X	X		X	X
Author-name Corporate	X			X	X
Date/Time Last Modified			X		X
Record Source	X			X	X
Distributor Name	X			X	X
Index Term—Controlled	X			X	X
Local Subject Index	X			X	X
Any	X			X	X

TABLE 2
GILS ABSTRACT RECORD STRUCTURE (PARTIAL)

<i>Tag Path</i>	<i>Element</i>	<i>Mandatory?</i>	<i>Repeatable?</i>
(1,10)	rank	N	N
(1,12)	url	N	N
(1,14)	local control number	Y	N
(1,16)	dateOfLastModification	Y	N
(2,1)	title	Y	N
(4,1)	controlIdentifier	Y	N
(2,6)	abstract	Y	N
(4,51)	purpose	Y	N
(4,52)	originator	Y	N
(4,53)	accessConstraints	Y	N
(4,53)/(4,25)	generalAccessConstraints	Y	N
(4, 53)/(4,26)	originatorDisseminatorControl	N	N
(4,53)/(4,27)	securityClassificationControl	N	N
(4,71)	spatialDomain	N	N
(4,71)/(4,91)	boundingCoordinates	N	N
(4,71)/(4,91)/(4,9)	westBoundingCoordinate	N	N
(4,71)/(4,91)/(4,10)	eastBoundingCoordinate	N	N
(4,71)/(4,91)/(4,11)	northBoundingCoordinate	N	N
(4,71)/(4,91)/(4,12)	southBoundingCoordinate	N	N
(4,71)/(4,92)	place	N	Y
(4,71)/(4,92)/(4,13)	placeKeyword	N	N
(4,71)/(4,92)/(4,14)	placeKeywordThesaurus	N	N

The GILS Profile defines four Element Set Names. When a client submits a request asking the server to return a record using an Element Set Name, the server returns records with a specific set of elements from the database record. The four GILS Profile Element Set Names and the elements contained in each are:

- Element Set Name “B” contains at least Title, Control Identifier, Originator, and Local Control Number

Primitive Elements

<u>Tag</u>	<u>Element</u>	<u>Recommended Datatype</u>
1	controlIdentifier	InternationalString
9	westBoundingCoordinate	intUnit
10	eastBoundingCoordinate	intUnit
11	northBoundingCoordinate	intUnit
12	southBoundingCoordinate	intUnit
13	placeKeyword	InternationalString
14	placeKeywordThesaurus	InternationalString
25	generalAccessConstraints	InternationalString
26	originatorDisseminatorControl	InternationalString
27	securityClassificationControl	InternationalString

Constructed Elements

<u>Tag</u>	<u>Element</u>	
51	purpose	Constructed as follows— This element may include the element wellKnown and may also include locally defined elements.
52	originator	Constructed as follows— This element may include the element wellKnown and may also include locally defined elements.
53	accessConstraints	Constructed as follows— This element may include any of the following as well as locally defined elements: generalAccessConstraints, originatorDisseminatorControl, securityClassificationControl.
71	spatialDomain	Constructed as follows— This element may include any of the following as well as locally defined elements: boundingCoordinates, place.
91	boundingCoordinates	Constructed as follows— This element may include any of the following as well as locally defined elements: westBoundingCoordinate, eastBoundingCoordinate, northBoundingCoordinate, southBoundingCoordinate.
92	place	Constructed as follows— This element may include any of the following as well as locally defined elements: placeKeyword, placeKeywordThesaurus

Figure 4. GILS tagSet (Partial)

- Element Set Name “G” contains all B Element Set elements and Cross Reference
- Element Set Name “W” contains all B Element Set elements and bodyOfDisplay
- Element Set Name “F” contains all elements available in the record.

A GILS server should return to the client all of the elements specified by the Element Set Name if there are data available in the database record. In some cases it may not be possible to encode certain data in the requested record syntax (e.g., some types of locally defined binary data may

- SPATIAL DOMAIN** (Optional, Not Repeatable): This element is a grouping of subelements that together provide the geographic areal domain of the data set or information resource. Geographic names and coordinates can be used to define the bounds of coverage. Although described here informally, the spatial object constructs should be as defined in FIPS 173, "Spatial Data Transfer Standard."
- BOUNDING COORDINATES** (Optional, Not Repeatable): This subelement limits the coverage of a data set expressed by latitude and longitude values in the order western-most, eastern-most, northern-most, and southern-most. For data sets that include a complete band of latitude around the earth, the West Bounding Coordinate shall be assigned the value: -180.0, and the East Bounding Coordinate shall be assigned the value: 180.0. The following subelements comprise the Bounding Coordinates:
- WEST BOUNDING COORDINATE:** Western-most coordinate of the limit of the coverage expressed in longitude. Domain: -180.0 <= West Bounding Coordinate <= 180.0
 - EAST BOUNDING COORDINATE:** Eastern-most coordinate of the limit of coverage expressed in longitude. Domain: -180.0 <= East Bounding Coordinate <= 180.0
 - NORTH BOUNDING COORDINATE:** Northern-most coordinate of the limit of coverage expressed in latitude. Domain: -90.0 <= North Bounding Coordinate <= 90.0; North Bounding Coordinate >= South Bounding Coordinate
 - SOUTH BOUNDING COORDINATE:** Southern-most coordinate of the limit of coverage expressed in latitude. Domain: -90.0 <= South Bounding Coordinate <= 90.0; South Bounding Coordinate <= North Bounding Coordinate
- PLACE** (Optional, Repeatable): This subelement identifies geographic locations characterized by the data set or information resource through two associate constructs:
- PLACE KEYWORD:** The geographic name of a location covered by a dataset or information resource.
 - PLACE KEYWORD THESAURUS:** The name of a formally registered thesaurus or similar authoritative source of Place Keywords.

Figure 5. Spatial MetaData in GILS locator records

not be encodable in a USMARC or SUTRS record). In those cases, the data in those elements will not be returned.

A record syntax is a way of representing the retrieval record to return from the server to the client. Record syntaxes allow a server to return GILS Locator Records to a client in a format that the client can process. The GILS Profile requires a GILS server to support three record syntaxes: (1) Generic Record Syntax (GRS-1), (2) Simple Unstructured Text Record Syntax (SUTRS), and (3) USMARC. GRS-1 representation is considered the complete and canonical form since GRS-1 can use the structure provided in the GILS Schema to identify and represent both the well-known and locally identified elements.

SUTRS allows relatively simple clients to accept and display GILS Locator Records to users. SUTRS treats the GILS Locator Record as a block of text with no structure or content designation, and the client should not expect to be able to parse the record to obtain any individual GILS

elements. The client simply presents the record as it is provided by the server. The GILS Profile does suggest a preferred display format for use with SUTRS, but the display format is a concern only of the server and not the client.

USMARC is required since the library community will be a primary user community for GILS. The GILS Profile provides a mapping between the GILS Core Elements and USMARC for use with the USMARC record syntax.

The foregoing has described the basic components of the GILS Profile. Understanding the profile requires a basic understanding of Z39.50, how it works, and the terminology used in the standard.

GILS AND SPATIAL INFORMATION RESOURCES

An important set of information resources created, collected, or held by federal agencies relate to spatial data. Like other federal information resources, spatial information resources can be described by GILS Locator Records. GILS will be a useful mechanism for discovering spatial information resources throughout the federal government. Many federal agencies create or collect spatial data. The membership of the Federal Geographic Data Committee (FGDC) indicates the range of agencies involved with spatial data. Each of the Executive Branch agencies that are covered by OMB Bulletin 95-01 are required to create GILS Locator Records for their information resources including those related to spatial data. GILS will enable users to search across these agencies' Locator Records to discover and locate spatial information resources.

GILS will assist in achieving the goals of OMB Circular A-16 (1990), *Coordination of Surveying, Mapping, and Related Spatial Data Activities*. Those goals include the reduction of duplication and reduction of expense of developing geographic data. For example, a federal agency could use GILS to discover if there are existing spatial information resources collected by other agencies before initiating new projects that might duplicate existing data sets.

Gluck (1994) identifies three primary elements that characterize spatial data selection:

- *Attributes*, which define the contents and characteristics of spatial data
- *Time*, which can provide a time-scale of the coverage of the spatial data
- *User task*, which determines the appropriateness of components of a particular data set or other information resource to solve a particular problem (p. 640).

GILS Locator Records begin to address these selection criteria since metadata about spatial information resources are accommodated by spe-

cific GILS Core Elements. Figure 5 gives the GILS Core Elements that provide spatial metadata along with their semantics as found in the GILS Profile (note that the names and semantics of these elements reflect changes in a revised version of the GILS Profile currently under development as of April 1995).

Appendix A contains two examples of GILS Locator Records that describe spatial information resources. These records, created by the United States Geological Survey, describe:

- National Digital Cartographic Data Base—Large Scale (Record #1)
- Landsat Earth Resources Observations Multispectral Scanners Imagery (Record #2).

The following discussion refers to these records to indicate how a user might search, retrieve, and use these records.

GILS provides the possibility for users to search Locator Records using the spatial metadata elements. The GILS Profile identifies Use Attributes for the GILS Core Elements listed in Figure 5. As one hypothetical example, these Use Attributes could be used to retrieve Record #1 in Appendix A. A user would submit a query to a GILS server with one or more of the Use Attributes West Bounding Coordinate, East Bounding Coordinate, North Bounding Coordinate, South Bounding Coordinate with terms of -179, -66, 72, 24 respectively and a Relation Attribute of “equal” to retrieve the example Record #1. A more likely scenario, however, would be for a user to submit a query with these Use Attributes and a range of bounding coordinates to find GILS Locator Records that describe resources that have a spatial coverage the user is interested in.

It is important to realize, however, that these spatial metadata Use Attributes are not included in the minimum set of Use Attributes required by the GILS Profile. Therefore, this specific searching capability on spatial data fields may or may not be offered by all GILS conformant servers. There is nothing in the GILS Profile, however, that precludes agencies or other organizations from implementing a GILS server that provides robust searching of the spatial metadata elements in GILS Locator Records. An intermediary user of GILS could add value to GILS Locator Records by mounting them on an information system that provides specialized spatial data searching capabilities (i.e., ensures that the SPATIAL DOMAIN elements are access points for the Locator Records).

Record #2 in Appendix A describes a spatial information resource that is accessible via the World Wide Web protocol. This points to another important feature of GILS. If a user retrieves this GILS Locator Record and is interested in accessing the spatial information resource described

in the Locator Record, it is possible to use information in the record to automatically link to the resource. Under the AVAILABILITY element, there are two subelements, LINKAGE and LINKAGE TYPE. Record #2 contains a Uniform Resources Locator (URL) in the LINKAGE element (i.e., <http://sunl.cr.usgs.gov:80/glis/glis.html>). Appropriate client software would be able to use the LINKAGE information to start up a WWW browser and connect to the spatial information resource pointed to by the Locator Record. Such capabilities imply that the user's client will need to support multiple protocols in addition to Z39.50 (e.g., http, FTP, telnet, gopher). This scenario of seamless network navigation (i.e., discovery, searching, retrieval, and access) that GILS makes possible is a most exciting prospect.

A current initiative will extend the use of Z39.50 for information retrieval of spatial data and will work in concert with GILS. Similar to the GILS Profile, work is now underway to develop a Z39.50 application profile for the content specification of digital geospatial metadata. The GEO Profile development is being coordinated by Douglas Nebert at USGS in cooperation with ASTM Section D18.01.05 on Mapping and GIS for the FGDC (contact Nebert for additional information <ddnebert@usgs.gov>). The GEO Profile supports search and retrieval of geospatial metadata entries and related geospatial data sets accessible on GEO conformant servers. Implementations of the GEO Profile will provide Z39.50 access to existing or new data sets. The implication of this for GILS is that a Z39.50 client that supports both the GILS Profile and the GEO Profile will be able first to search GILS Locator Records and discover particular spatial information resources described by GILS Locator Records and then be able to establish a Z39.50 connection to the actual data sets and use Z39.50 functionality to do sophisticated searching on the data set. The GEO Profile is an important complement to the GILS.

GILS Core Elements provide metadata information for a broad range of information resources including spatial information resources. The spatial metadata elements in GILS Locator Records will assist users of spatial data to discover and select spatial resources that are appropriate to their information problems.

SUMMARY AND CONCLUSION

This paper has provided an introduction and overview to the GILS, the new federal initiative to improve access to, and management of, government information resources. Z39.50 is an important foundation for a decentralized, agency-based, and Internet-accessible GILS. The discussion of Z39.50 and how it is used in GILS demonstrates the utility of using an information retrieval protocol for the discovery of, and access to, federal information resources.

Federal agencies have begun to create GILS Locator Records and are making plans to install Z39.50 access to the information servers and databases that contain the Locator Records. The various user communities that have a stake in accessing federal information resources, especially spatial information resources, will want to encourage agencies to move ahead aggressively in establishing GILS servers and to provide robust searching and retrieval capabilities on those servers. Z39.50 is rich in functionality for searching and retrieving information, and any limits on how users will be able to search GILS Locator Records will likely be a function of the database management and search engines deployed by agencies. Users will want to state their needs to agencies, and press agencies especially for the capabilities of searching the spatial metadata elements in GILS Locator Records.

The OMB Bulletin 95-01 that establishes GILS sets a number of deadlines for agencies. By December 31, 1995, agencies are to have created their initial GILS Core Locator Records. By the same date, these records are to be available online in a form compliant with the GILS Profile. The success of this important policy initiative to provide better public access to government information will depend on the federal agencies complying with the OMB Bulletin. The spatial data user community is an important stakeholder and has the ability to press federal agencies to establish GILS servers so that the community can discover, identify, and locate the wealth of spatial information resources available through the federal government.

APPENDIX A

SAMPLE GILS RECORDS

This appendix contains two GILS Locator Records created by the United States Geological Survey. They describe spatial information resources.

RECORD #1

Title: NATIONAL DIGITAL CARTOGRAPHIC DATA BASE-LARGE SCALE

Acronym: NDCDB/DLG

Originator: USGS/NMD

Local-Subject-Index: AEDD; ALASKA; ALASKA DIRECTORY; ARCTIC; CARTOGRAPHY; DLG; DOIGC; GEODATA; MAP; USGS; ESDD; U.S. Federal GILS

Abstract: Contained are selected US regional coverage of planimetric map features and contours. Also, included are digitized source-map scales varying from 1:24,000 to 1:62,500. The largest amount of coverage exists for the Public Land Survey System (PLSS) and boundary data categories. Data are added on the basis of mapping requirements of the USGS and other Federal agencies. The information is stored in topologically structured DLG-3 format. This portion of the Digital Cartographic Data Base is referred to as US GeoData. The data are distributed through the National Mapping Division. Contact the Earth Science Information Center for documentation, availability of specific coverage, output formats, and current price. The data are not available for general online access.

Format: DIGITAL DATA SETS

Spatial-Domain:

Geographic-Coverage: UNITED STATES

Coordinate-System: UTM NORTHINGS AND EASTINGS; ARBITRARY X, Y COORDINATES

Coverage-Description: NONE REPORTED

Bounding-Coordinates:

West-Bounding-Coordinate: -179

East-Bounding-Coordinate: -66

North-Bounding-Coordinate: 72

South-Bounding-Coordinate: 24

Time-Period:

Time-Period-Textual: 1979-PRESENT

Availability:

Distributor:

Name: USGS/NMD

Organization: USGS/NMD

Street-Address: U.S. GEOLOGICAL SURVEY, RESTON-ESIC, 507 NATIONAL CENTER

City: RESTON

State: VA

Zip-Code: 22092

Country: USA

Telephone: (703)860-6045

MOEN/DISCOVERING, IDENTIFYING, & ACCESSING SPATIAL DATA

Resource-Description: NATIONAL DIGITAL CARTOGRAPHIC DATA BASE-LARGE SCALE

Order-Process: Contact any below-listed USGS Earth Science Information Center (ESIC) for assistance: Anchorage-ESIC, 4230 University Dr., Rm 101, Anchorage, AK 99508-4664 (907)786-7011; Anchorage-ESIC, U.S. Courthouse, Rm 113, 222 W. 7th Ave., #53, Anchorage, AK 99513-7546 (907)271-4307; Denver-ESIC, 169 Federal Bldg., 1961 Stout St., Denver, CO 80225-0046 (303)844-4169; Lakewood-ESIC, Box 25046, Federal Ctr., MS 504, Denver, CO 80225-0046 (303)236-5829; Rolla-ESIC, 1400 Independence Rd., MS 231, Rolla, MO 65401 (314)341-0851; Salt Lake City-ESIC, 8105 Federal Bldg., 125 S. State St., Salt Lake City, UT 84138 (801)524-5652; San Francisco-ESIC, 504 Custom House, 555 Battery St., San Francisco, CA 94111 (415)556-5627; Spokane-ESIC, 678 U.S. Courthouse, W. 920 Riverside Ave., Spokane, WA 99201 (509)353-2524; Stennis Space Ctr.-ESIC, Bldg. 3101, Stennis Space Center, MS 39529 (601)688-3544; Washington, D.C.-ESIC, Dept. of the Interior Bldg., 18th & C Sts., NW, Rm. 2650, Washington, D.C. 20240 (202)343-8073; Menlo Park-ESIC, Building 3, MS 532, 345 Middlefield Rd., Menlo Park, CA 94025 (415)329-4309; Reston-ESIC 507 National Center Reston, VA 22092 (703)860-6045

Technical-Prerequisites:

Data-Set-Type: AUTOMATED

Computer-Type: AMDAHL 5890

Computer-Location: RESTON, VA

Access-Constraints: Access is not restricted unless otherwise noted.

Documentation: Standards for Digital Line Graphs (6 parts), Br. OF Technical Mgmt., MS 510, USGS, Reston, VA. Digital Line Graphs from 1:24,000-scale Maps, Data Users Guide 1, 1986, USGS.

Use-Constraints: These data and information have been approved for release by the Director of the USGS on condition that neither the USGS nor the United States Government may be held liable for any damages resulting from its authorized or unauthorized use.

Status: OPERATIONAL

Point-of-Contact:

Organization: USGS/NMD

Street-Address: U.S. GEOLOGICAL SURVEY, RESTON-ESIC, 507 NATIONAL CENTER

City: RESTON

State: VA

Zip-Code: 22092

Country: USA

Telephone: (703)860-6045

Purpose: These data and information resources contribute to the USGS mission of Earth science in the public service and the USGS role as the principal source of earth-science research and information for the Nation.

Control-Identifier: USGS0024

Record-Source: USGS/NMD

Date-of-Last-Modification: 9008

RECORD #2

Title: LANDSAT EARTH RESOURCES OBSERVATIONS MULTISPECTRAL SCANNERS IMAGERY

Acronym: LANDSAT DATA (MSS)

Originator: USGS/NMD

Local-Subject-Index: DOIGC; EDC; EROS; GLOBAL; IMAGERY; LANDSAT; REMOTE SENSING; SATELLITE; US; USGS; ESDD; U.S. Federal GILS

Abstract: A Multispectral Scanner (MSS) has flown on board five Landsat satellites to date. Landsat 1, 2 and 3 operated in a circular, sun-synchronous, near-polar orbit at an altitude of approximately 913 km (567 miles), with a nominal 9:30 AM crossing of the Equator during the descending mode. They circled the Earth every 103 minutes, completing 14 orbits per day and viewing the entire earth every 18 days. The Landsat orbits are selected and trimmed so that each satellite ground trace repeats its Earth coverage at the same local time every day. Landsat 4 and 5 also operate in circular, sun-synchronous, near-polar orbit at an altitude of 705 km (438 miles), with a nominal 9:45 AM crossing at the Equator during the descending mode. Each orbit takes nearly 99 minutes, and the spacecrafts complete just over 14 orbits each day, covering the entire Earth (poles excluded) every 16 days. An international Landsat database exists that contains information on the available MSS data acquired by independent international ground receiving stations. These data are directly downlinked to ground receiving stations when in transmission range of the satellite(s). The framing of each scene is based on the World Wide Reference System (WRS), a network of intersecting paths and rows whose junctions define the nominal scene center of each Landsat scene. The WRS path represents the nominal satellite track, a maximum of 251 paths exist for Landsat 1-3 and 233 paths exist for Landsat 4 and 5 data. The WRS row indicator represents scene centers that are chosen at 23.92 second increments along the orbital track, a total of 248 row numbers exist. MSS imagery is available on microimage fiche, otherwise in late 1994, MSS browse imagery is scheduled to be available on GILS. This MSS browse imagery may be viewed via the EDC World Wide Web home page, which allows direct access to GLIS. To date, sixteen unique ground stations have acquired and archived MSS data on a wide variety of media. EROS Data Center periodically receives international database tape catalogs of foreign stations' data holdings which are added into an on-line Landsat global database (i.e., International Landsat Database).

Format: MICROFILM; PHOTOGRAPHS; DIGITAL DATA SETS AND BROWSE

Spatial-Domain:

Geographic-Coverage: GLOBAL

Coverage-Description: NONE REPORTED

Time-Period:

Time-Period-Textual: 1972-1993

Availability:

Distributor:

Name: USGS/NMD

Organization: USGS/NMD

Street-Address: EROS DATA CENTER, U.S. GEOLOGICAL SURVEY

City: SIOUX FALLS

State: SD

Zip-Code: 57198

MOEN/DISCOVERING, IDENTIFYING, & ACCESSING SPATIAL DATA

Country: USA

Telephone: (605)594-6151

Fax: (605)594-6589

Resource-Description: LANDSAT EARTH RESOURCES OBSERVATIONS
MULTISPECTRAL SCANNERS IMAGERY

Order-Process: To place orders, obtain additional information, technical details, ancillary products, and pricing schedules regarding products and services, or international data holdings contact the EROS Data Center, Customer Services Section.

Technical-Prerequisites:

Data-Set-Type: AUTOMATED

Access-Method: INTERACTIVE AND BATCH

Number-of-Records: 650,000

Bytes-Per-Record: 154

Computer-Type: IBM

Computer-Location: SIOUX FALLS, SD

Linkage: <http://sunl.cr.usgs.gov:80/glis/glis.html>

Linkage-Type: URL

Access-Constraints: Access is not restricted unless otherwise noted.

Documentation: U.S. Geological Survey, 1979, Landsat Data Users Handbook REVISED): USGS, p. 1-1 TO AH-1. U.S. Geological Survey and National Oceanic and Atmospheric Admin., 1984, Landsat, 4 handbooks.

Use-Constraints: These data and information have been approved for release by the Director of the USGS on condition that neither the USGS nor the United States Government may be held liable for any damages resulting from its authorized or unauthorized use.

Status: OPERATIONAL

Point-of-Contact:

Name: CUSTOMER SERVICES

Organization: USGS/NMD

Street-Address: EROS DATA CENTER, U.S. GEOLOGICAL SURVEY

City: SIOUX FALLS

State: SD

Zip-Code: 57198

Country: USA

Telephone: (605)594-6151

Fax: (605)594-6589

Purpose: These data and information resources contribute to the USGS mission of earth science in the public service and the USGS role as the principal source of earth-science research and information for the Nation.

Control-Identifier: USGS2020

Record-Source: USGS/NMD

Date-of-Last-Modification: 9410

REFERENCES

- Christian, E. (1994). *Government information locator service (GILS): Report to the Information Infrastructure Task Force*. Available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at <130.11.48.107> as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils/gils.txt (ASCII text format).
- Gluck, M. (1994). Spatial information and information science: Introduction to JASIS' special topic issue on spatial information. *Journal of the American Society for Information Science*, 45(9), 640-644.
- Information Infrastructure Task Force. (1993). *The national information infrastructure: Agenda for action*. Washington, DC: NTIA NII Office, Department of Commerce. Available in ASCII text format, filename: niiagenda.asc on the NTIA Bulletin Board (202-482-1199) and the Fedworld bulletin board (703-321-8020); available by anonymous FTP from <ftp.ntia.doc.gov> as /pub/niiagenda.asc
- International Organization for Standardization/International Electrotechnical Commission. (1992). *ISO/IEC TR10000-1 Information technology—Framework and taxonomy of international standardized profiles—Part 1: Framework*. Geneva: ISO/IEC Copyright Office.
- Lynch, C. A. (1992). Online searching on the internet: The challenge of information semantics for networked information. In M. E. Williams (Ed.), *Proceedings of the 1992 National Online Meeting* (pp. 203-210). Medford, NJ: Learned Information.
- McClure, C. R.; Bishop, A.; Doty, P.; & Bergeron, P. (1990). *Federal information inventory/locator systems: From burden to benefit: Final report*. Syracuse, NY: Syracuse University, School of Information Studies. (ERIC Document Reproduction Service Number ED 326 247)
- McClure, C. R.; Ryan, J.; & Moen, W. E. (1992). *Identifying and describing federal information inventory/locator systems: Design for networked-based locators*. Capitol Heights, MD: National Audio Visual Center. (ERIC Document Reproduction Service Number ED 349 031)
- Michael, J. J., & Hinnebusch, M. (1995). *From A to Z39.50: A networking primer*. Westport, CT: Mecklermedia.
- Moen, W. E., & McClure, C. R. (1994). *The government information locator service (GILS): Expanding research and development on the ANSI/NISO Z39.50 information retrieval standard*. Bethesda, MD: NISO Press. Available via FTP at <ftp.cni.org> as pub/gils/profile/final.report/gils—report.txt
- National Information Standards Organization. (1992). *ANSI/NISO Z39.50-1992. American national standard information retrieval application service definition and protocol specification for open systems interconnection*. Gaithersburg, MD: NISO Press.
- National Archives and Records Administration. (1995). *Guidelines for the preparation of GILS core entries*. Available electronically as <URL: <http://gopher.word.gov:70/1/managers/gils/guidance>>
- National Information Standards Organization. (1995). *ANSI/NISO Z39.50-1992, American national standard information retrieval application service definition and protocol specification*. [Post-ballot Draft]. Available via FTP<ftp.loc.gov> as pub/z3950/v3.ballot+
- National Institute of Standards and Technology, Computer Systems Laboratory. (1994). *Application profile for the government information locator service (GILS)*, (FIPS Pub 192). Gaithersburg, MD: U.S. Department of Commerce (National Institute of Standards and Technology, Computer Systems Laboratory).
- National Performance Review. Office of the Vice President. (1993). *Reengineering through information technology: Accompanying report of the national performance review*. Washington, DC: USGPO.
- Office of Management and Budget. (1990). *Coordination of surveying, mapping, and related spatial data activities* [Circular No. A-16]. Washington, DC: Office of Management and Budget.
- Office of Management and Budget. (1994a). *Establishment of Government Information Locator Service*. [Bulletin No. 95-01]. Washington, DC: Office of Management and Budget.

MOEN/DISCOVERING, IDENTIFYING, & ACCESSING SPATIAL DATA

Office of Management and Budget. (1994b). Management of federal information resources. *Federal Register*, 59, 37906-37928.

BARBARA P. BUTTENFIELD

Geographic Information Systems and Digital Libraries: Issues of Size and Scalability

The term "scalability" has specific connotations in Geographic Information Systems (GIS) that conventionally relate to monitoring and predicting growth of geographic phenomena. A family of computational models has been developed to predict changes in structure associated with changes in size. These models have been applied in physical science, social science, and cartographic science to study growth and may assist in monitoring the growth of digital libraries as well. As the size of the digital library increases, challenges for data organization and collection maintenance tasks will also increase. However, the rate of increase may not be in linear proportion to library size. At some critical scales, existing procedures will fail and new procedures must be implemented to accommodate further growth. Allometric principles may be applied to estimate these critical scales. Three aspects (data volume, indexing, and metadata recordation) will be discussed in the context of implementing and maintaining a digital library containing spatial data archives distributed across local or global electronic networks.

THE SCALE OF GEOGRAPHIC INFORMATION

The nature of spatial data is that phenomena and processes take on varying appearances with variations in resolution at which they are measured and observed. For example, a high altitude, 79-meter resolution satellite image of a large city may display land-water distinctions and areas of vegetative cover distinguished from paved or built-up areas, while a lower altitude, 10-meter aerial photograph can be used to identify urban land use patterns and large landmarks. Geographic processes that are evident at small scales include geologic processes such as continental plate tectonics or global migration patterns. At larger scales, processes such as erosion and urban zoning constraints become evident. Environmental and atmospheric scientists are sensitive to the scales at which a phenomenon or process can be identified in a map or satellite image and choose their data sources accordingly.

To meet the needs of the environmental and atmospheric scientists, cartographers must preserve details that are necessary to identify geographic phenomena and processes when they change the scale of map data. Many think that changing map scale involves either enlarging or

eliminating detail. This view is incorrect. The cartographic challenge is that new details are needed to preserve realism at larger scales. As one "zooms" in or out, geographic details may appear, then disappear, and sometimes reappear. Some feature domains (hydrography) change rapidly with changes in map scale, while others (transportation) stabilize at a particular level of detail. The scales at which changes occur differ from one domain to another and are not straightforward to predict due to variations in terrain, soil type, and other mediating factors. Much of the focus in current cartographic research centers on formalizing knowledge about what geographic details can be identified at particular data scales, and on developing computational and graphic procedures to meet user requirements for information whose appearance may change with scale. The challenge to our discipline is to create and maintain digital cartographic data sets that embody multiple representations of features and attributes, and to identify the ranges of scale for which data processing algorithms are effective. At the limits of these ranges, operating procedures must be changed to preserve geographic and visual logic in the database.

MEASURING THE SIZE AND SCALE OF LIBRARIES

It is intriguing for a cartographer to discover that library scientists face a similar challenge. The challenges in creating and maintaining a digital library include issues of interface design (Siegel, 1991), adaptation of existing procedures (Cohn et al., 1992; Weibel, 1992), and the new roles for library and information scientists arising with emerging technologies (Smith & Dalrymple, 1992). The challenge relating to size and scale is that as the library expands and matures, much of the collection maintenance must be modified to meet information demands and user expectations. In some cases, the larger size will require altering the contents of existing archives. It is likely that as more information becomes available, it becomes more difficult to access, retrieve, and catalog. This paradox becomes especially evident for digital libraries that are electronically networked to other distributed archives. The effectiveness of digital libraries will be based in part upon the ability of library scientists to "scale up" operating procedures. Scaling up cannot be accomplished by mimicry of existing procedures and often involves evolutionary or revolutionary change. For example, conversion from the Dewey Decimal cataloging scheme to Library of Congress Cuttering scheme in response to the increase in library holdings at sites across the country also introduced new methods of formal specification describing the contents of archived items (Molz, 1984).

This discussion will expand upon concepts of scale and growth as they are applied in a variety of natural and social science disciplines and present

models by which scale progressions and their consequences may be anticipated. The objective of such study in other disciplines can facilitate information gathering or data analysis, or estimate consequences of predicted growth on large or complex phenomena, or improve management by foreseeing points of growth at which operating procedures must change. The scaling concepts are referred to in other disciplines as allometry, which is the study of changes in size that are accompanied by changes in form or structure. Allometry has been applied to biological evolution (Gould, 1966), architectural engineering (Bon, 1973), industrial management (Haire, 1973), urban planning (Woldenberg, 1971, 1973), and in many other disciplines. There is a long history of allometric modeling in GIS which began with early cartographic studies by Richardson (1961) and Mandelbrot (1967). Their work demonstrated that the length of coastlines and other map features tends to increase without apparent limit with finer units of measurement (see Figure 1). A recent survey by Lam and DeCola (1993) demonstrates the breadth of recent applications in GIS and geography.

Allometric study should also prove useful for management of digital libraries, particularly as it becomes clear that electronic information depositories will continue to come online and to grow even though it is currently difficult to predict just how large and intertwined these may become. Kemeny's (1962) projections of exponential library growth accept that holdings may be distributed in multiple branches, but his assumption that the library items take up physical space have been surpassed with the advent of electronic archival dissemination. A real challenge for those who monitor the growth of digital libraries lies in determining how to measure the size of holdings at any point in time.

CONCEPTS OF SCALE AND SCALE CHANGE

For the purpose of this presentation and the discussion that it may generate, let the reader accept the label "scalability" to encompass the range of issues associated with changing relative or absolute scale. Scalability takes the same linguistic root as the word scale. In cartography, scale is the ratio between distance on a map and distance on the earth, and customarily reported as a Representative Fraction (RF). To belabor the point, an RF value of 1:24,000 describes a map where one map unit equals 24,000 earth units or 1 inch to 2,000 feet. A more general definition implying increases of scale denotes "a succession or progression of steps or degrees; a graduated series (the scale of taxation, a social scale) or a point on such a scale" (*Random House Unabridged Dictionary of the English Language*). The last clause of the definition ("a point on such a scale") leads in many cases to considering scale and size interchangeably. A variety of

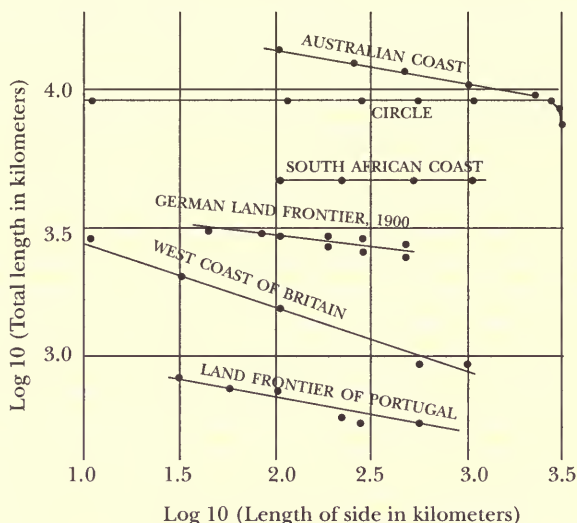


Figure 1. Increase in the length of cartographic lines measured at finer and finer resolution. (Redrawn from Richardson, 1961)

attributes are applied to quantify scale progressions. It is possible to create unusual cartographic transformations by intermingling scale units, as for example in scaling the size of countries in proportion to energy generation or consumption.

Scale measures may be applied to understand the size and growth of libraries and of digital libraries in particular. Numerous units of measurement might be formulated, for example, the number of items (books, journal series, map series) acquired per year, the number of items contained in these items (number of chapters, journal articles, or map sheets, respectively), the number of patrons, the number of branches if the library is distributed, etc. RF values might be developed to compare the acquisitions per library branch, the average number of items requested per patron per year, or the number of library staff in relation to the number of items cataloged. Clearly it is not the place for a cartographer to tell a librarian how to determine what measures of library size are appropriate. Likewise, telling the library community that digital library archives will tend to grow without apparent limit is a form of "preaching to the choir." The intention of this discussion is to outline quantifiable methods which may assist in estimating rates of digital library expansion, and adjusting operational procedures to accommodate increasing acquisitions and item retrieval.

A very powerful approach has been used in other scientific disciplines to model changes in scale. As it turns out, changes in gross structure (for example overall size) are in proportion with changes in substructures. To take a simple example, while the perimeter of a circle ($2\pi r$) increases linearly, its area (πr^2) increases geometrically. For very large circles, as the great and small circles of latitude circling the globe, the difference between the coefficients π and 2π becomes inconsequential with respect to the difference between r and r^2 , and thus scalability is a factor of the radius alone. This knowledge enables cartographers to estimate linear distances on the sphere which vary in direct proportion to the cosine of latitude. What is important here is not the specific formula, but that these estimations can be made, and made much more simply than actually going to a place and surveying distances. In biology, the focus of attention is more often placed upon animal weight and animal height (or length). So, for example, a fish may be seen to double its weight in growing from four to five inches long, and knowing this a biologist may predict the weight of a very large fish simply by measuring its length and applying the scale ratio $W=kL^3$ (Thompson, 1977, p. 16). In this way, potential fishing yields can be estimated by measuring the length of schooling fish on air photographs.

The accuracy of such estimation is based of course on the assumption that the shape of the globe, or the shape of the fish at both sizes is equivalent or isometric. This is called the Principle of Similitude and can be applied to estimate multiscaled processes such as cohesion, chemical, electrical, and gravitational attraction at molecular and astronomical scales. The principle sounds pretty simple, and for geometric objects (like circles) it is. However, isometric relationships occurring in nature tend to hold true only within finite ranges of scale. At certain critical scales, the isometric model fails to generate an accurate estimate. Growth beyond these critical scales is associated with changes in form and proportion that are called allometric. Thompson (1977) demonstrates the principle for an engineering example:

the strength of an iron girder obviously varies with the cross-section of its members, and each cross-section varies as the square of a linear dimension; but the weight of the whole structure varies as the cube of its linear dimensions. It follows at once that, if we build two bridges geometrically similar, the larger is the weaker of the two, and is so in the ratio of their linear dimensions. (p. 18)

And later in the passage, Thompson (1977) refers to Galileo's writings in the fifteenth century, and a comment that when building things at increasing scale, eventually

beams and bolts would cease to hold together; nor can Nature grow a tree nor construct an animal beyond a certain size, while retaining the proportions and employing the materials which suffice in the case of a smaller structure. The thing will fall to pieces of its own weight unless we either change its relative proportions, which will cause it to become clumsy, monstrous or inefficient, or else we must find new material, harder and stronger than was used before. (p. 19)

Allometric models apply numeric power laws relating internal factors (growth, weight, or mass) acting within an organism to the external forces (environmental, gravitational, etc.) acting upon it. The approach focuses on apparently paradoxical changes in structure that accompany growth, evolution, and maturation. The premise in allometry is that as critical size thresholds are passed at certain points in the growth process, internal physiological changes take place to accommodate the increased external forces acting on the organism. Allometry describes changes in shape and form that accommodate necessary changes in physiology that occur with scale change. Inversely, by identifying the critical points where changes in form occur, one can predict when, in the growth process, the internal or physiological changes ought to take place.

In physics, chemistry, architecture, and so on, the "materials" referred to in the quotes above are physical materials such as gypsum, wood, and steel. In social and informational sciences (cartography, communication, library science), the "material" may be considered as a metaphorical reference to information content or structure. New structure is generated by reshaping the existing structure, whether through changes in organizational operating procedures, or selection of different computational algorithms. The parameters guiding the algorithms modify the form of the organization or the structure of the information.

The scale at which size change becomes allometric can be identified where the ratio changes between the subcomponent parts and the whole. If the rate of change exceeds the proportion for isometric conditions, one speaks of positive allometry. Otherwise, the allometric relationship is said to be negative. Allometries may be quantified: when plotted on a graph, isometric relations will display values in proportion one would expect given the scale ratio. Signed allometries will have slopes that are greater than or less than this expected value respectively (see Figure 2). In the Euclidean cross-tabulation mentioned above, each cell in the matrix is characterized by allometric relations of a fixed magnitude. Comparisons between volume allow extensions from consideration of a single organism to consider the growth of organizations. For example, Haire (1973) studied industrial firms, comparing the number of employees who dealt primarily with activities inside or outside the firm. The internal staff include personnel officers, for example, while the external staff are in marketing and purchasing. Haire (1973) argued:

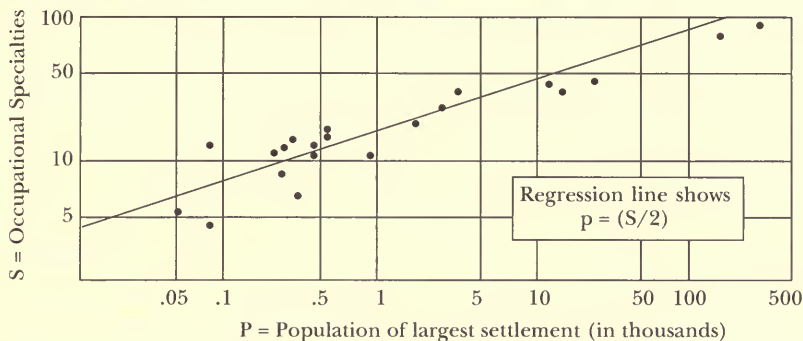


Figure 2. Relation between total number of occupational specialties and population of largest tribal settlement. (Redrawn from Naroll and Bertalanffy, 1956)

As the organization grows, its internal shape must change. Additional functions of coordination, control and communication must be provided and supported by the same kind of force that previously supported an organization without these things...If each increment in size produced one increment of additional supportive function, there would be no limit. However, in the organism, the proportion of skeleton needed to support the mass grows faster than the mass itself and...hence comes to consume a disproportionate amount of the productive capacity of the organization. It becomes important to identify the skeletal support of the firm, the forces it resists, and the rates at which the support must grow. (p. 264)

Haire's data are shown in Figure 3. He plots the cube root of internal staff (assuming that the "mass" of an organization should increase volumetrically) against the square root of the external staff (assuming these to represent the "skeletal" support structure), and finds nearly linear relationships in every case. He remarks: "There is no immediately obvious organizational artifact that imposes this orderly progression on each of the growth patterns. It looks as if a geometry very similar to conventional spatial description can be used in picturing social bodies" (Haire, 1973, p. 265).

"SCALING UP" DIGITAL LIBRARIES

It is clear from the literature that allometric models can be used to formalize estimations of size and growth, and these models can inform organizational planning and policy. How can allometric models be generated that describe the growth of digital libraries? Three areas come to mind where measures of scale might be applied to monitor growth and change: data volume, indexing, and metadata recordation. The three

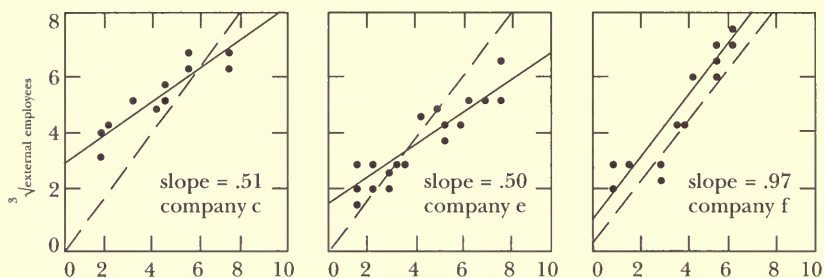


Figure 3. Allometry of Growth in Three Industrial Firms (dashed line indicates one-to-one ratio). Redrawn from Haire, 1973, p. 265.

areas are illuminated in the context of spatial data, which is the information customarily found within a Geographic Information System. GIS data are commonly troublesome for library archives in these three respects at least. To avoid these troublesome issues, most libraries separate regular (text) archives from (map and image) special collections. A truly workable digital library should eliminate distinctions between text and special archives (e.g., graphics, sound, video) or at least make such distinctions transparent to users. As the size of the digital library increases (this is the internal force or mass of the library), these issues are likely to become weak points in the supporting skeletal structure. Without proper adjustment, the sheer “weight” of the digital library will stress the overall library structure to the point where access and retrieval functions are likely to collapse.

Data Volume

As the number of items archived in a digital library increase, one can expect increasing strain on mass storage space, on data structures that organize the archives, and on the length of time required to retrieve any single item. One operational adjustment that would reduce overall storage volume is data compression. Storing maps and imagery in compressed form will reduce storage needs, and may shorten the time required for electronic data delivery. Many types of compression algorithms are designed to compress specific types of data, including static imagery (JPEG compression) or animated sequences (MPEG compression). The efficiency of the compression algorithm (total reduction in file size) is offset by consideration of how much information is lost when the file is reconstructed. Wavelet decomposition provides a hierarchical compression method

characterized by little or no loss of original information, and this method should apply well for reducing the volume of GIS archives. The disadvantage of archiving data in compressed form is that, at present, it is not possible to search the content of a compressed file. One must decompress a satellite image file to determine its geographical content. If images could be searched prior to decompression, query response times would be reduced, especially if data archives were distributed across a network.

Indexing

For geographical data in particular, it is advantageous to organize maps or files into series. Journal and monograph series run in linear sequence (volume 1, volume 2, etc.). Map series require two or more dimensions for proper organization (the third dimension is needed to archive duplicate map sheets and/or multiple editions). Organizing maps into spatial files in the digital archive shortens access times for patron requests for information that is contained in abutting map sheets. Digital indexing schemes for efficient searching of a two-dimensional array include Peano curves, Quadrees, and Morton indexing. Interface designs to facilitate user views on an archived map series are easy to design for single map series. Because patrons do not customarily request a series sheet name or reference number, but instead base their request on geographical place names or features, the user interface design must also display selected classes of geographic content in addition to the series footprints. However, as the digital archives incorporate additional series covering the same region, the user interface can become cluttered and unusable. These design issues must be considered before the library grows, and GIS technology (map overlay, feature buffering and selection) can be applied to assist the user in decomplicating interface designs for map indexing.

Metadata Recordation

Metadata are information about data. It includes, but is not limited to, the information customarily included in a MARC record. For digital geographical data, as of this calendar year, federal agencies distributing and exchanging spatial data are mandated to include metadata reports. The content of the metadata report is standardized first to determine what data exist, and second to determine fitness for use. Measures of data quality form the basis for determining fitness for use, which can be determined by empirical or deductive testing. Data quality measures include evaluation of positional and attribute accuracy, data completeness and currentness, and the logical consistency of underlying data structures. Third, as a digital library grows, and as its files increase in size, patrons should be able to browse a file's metadata records to learn in what data

format, data structure, and data model a data set will be delivered. In the optimal setting, patrons should understand the consequences of requesting an actual dataset before they download it in the event it is especially large (satellite image sets can run into hundreds of megabytes in size, and these archives grow larger every year) or complex. Fourth, metadata should report a chronology of data processing steps, that is, a lineage of the data since its original distribution. For GIS data in particular, lineage information provides a real challenge to record data changes including data filtering, changes in projection, category aggregation or reselection, and so on. Should every processing change be recorded, one might anticipate metadata reports containing unlimited record lengths—plainly this is unrealistic, especially for data delivery over a network. The question of how to transport and exchange metadata must be resolved quickly since metadata records (and particularly lineage information) will grow with data use, regardless of the growth of the digital library as a whole. Patrons' metadata browsing behaviors are not well understood, which provides another area that needs research. What may become necessary is the implementation of query mechanisms for searches that have not been anticipated by system designers.

SUMMARY

This discussion presents issues of scale as a cartographer views them and reviews how a class of models geared toward the description and analysis of scale change have been applied in several disciplines, including natural and social science. Classes of allometric relations can be defined taxonomically in a matrix cross-tabulating one-, two-, and three-dimensional phenomena in Euclidean space. This presentation takes examples from a few of the possible combinations. One class describes the growth and size of linear phenomena (such as the length of coastlines) as often applied in cartography. A second class of models identifies a square-cube relationship in which external growth of an organism or an organization will falter due to weakened internal support structures that must be modified if external growth is to continue.

Allometric models appear to have relevance for adoption of digital library technology, particularly as it becomes clear that electronic information repositories will continue to come online and to grow even though it is currently difficult to predict just how large and intertwined they may become. Quantifiable examples are more appropriately formulated by the library and information science community, since these are the individuals who currently measure and monitor library size and growth rates. The discussion has covered three areas of growth for which a digital library of GIS information will become especially challenged during intensive growth phases. These are presented as examples to guide discussion

and are not intended as an exhaustive list. It is up to the library community to select quantitative parameters for the models and to interpret them.

In closing, it is helpful to consider the perspectives from geography and library and information science alike.

Allometric concepts and the analysis of relative growth may help to fill our current vacuum of ignorance concerning relevant norms of societal growth and change. As humanity takes conscious control of the planet which shaped the species, the analysis of relative growth can indicate what changes are possible, which are most likely, and to some degree, which may be desirable. By attending to changes in shape of our social organism, we may become more competent in shaping change. (Dutton, 1973, p. 306)

“Embedded in the public library movement is the belief that people have a fundamental right to know. No matter how rich you are, how old you are, where you come from, or where you call home, you have a right to both information and knowledge” (Bremmer, 1994, p. 1). In addressing issues of digital library growth, there is an implicit requirement to attend to the needs of both library patrons and of library staff, meaning that access to archived items must serve multiple purposes, and information delivery must be flexible with respect to both content and presentation. One can assume that user needs will change with changes in the scale of the library. In this regard, the target “user community” and the set of information requirements is somewhat more complex than for the cartographic situation posed at the beginning of this discussion.

There are other issues germane to scalability of digital libraries, including issues of copyright and intellectual freedom, issues of privacy, issues of equality of access, and economic factors. Then, too, there is the role of the librarian in the digital library. In some circles, the advent of digital libraries is seen as a threat to the job security of library staff. Nothing could be further from the truth—the role of library and information science has never been so important as it is now.

I believe it is time we take much more seriously the important responsibility we hold in adopting the technologies now rolling out of Silicon Valley workshops. We need to evaluate them carefully before we buy [into] them. We need to make others aware of potential problems we see before others buy them. We urgently need “environmental impact studies” for new information technologies, so as to protect those good parts of our world information environment—like scholarly journals and neighborhood newspapers—that are on the “endangered species” list. Above all, we need to learn more about economics, and learn fast. (Nielson, 1981, p. 112)

ACKNOWLEDGMENTS

This paper forms a portion of the Alexandria Project, funded jointly by NSF, ARPA, and NASA (NSF grant IRI-94-11330). The research is also related to

Research Initiative 8 "Formalizing Cartographic Knowledge," a part of the National Center for Geographic Information and Analysis (NCGIA) (NSF grant SBE-88-10917). Support from these funding agencies is acknowledged. Comments on early versions by my colleagues Geoff Dutton and Michael Woldenberg have improved the logic and content considerably, and I am grateful for their advice and debate.

REFERENCES

- Bon, R. (1973). Allometry in the topologic structure of architectural spatial systems. *Ekistics*, 36(215), 270-276.
- Bremmer, S. W. (1994). *Long range planning* (How-to-do-it manual for librarians series, Number 40). New York: Neal-Schuman Publishers.
- Cohn, J. M.; Kelsey, A. L.; & Fiels, K. M. (1992). *Planning for automation* (How-to-Do-it Manual for Librarians Series, Number 25). New York: Neal-Schuman Publishers.
- Dutton, G. (1973). Criteria of growth in urban systems. *Ekistics*, 36(215), 298-306.
- Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews*, 41(4), 587-640.
- Haire, M. (1973). Biological models and empirical histories of the growth of organizations. *Ekistics*, 36(215), 263-269.
- Kemeny, J. G. (1962). A library for 2000 A.D. In M. Greenberger (Ed.), *Computers and the world of the future* (pp. 135-162). Cambridge, MA: MIT Press.
- Lam, N. S., & DeCola, L. (Eds.). (1993). *Fractals in geography*. Englewood Cliffs, NJ: Prentice-Hall.
- Mandelbrot, B. B. (1967). How long is the coast of Britain? Statistical self-similarity and fractal dimension. *Science*, 156(May 5), 636-638.
- Molz, R. K. (1984). *National planning for library service, 1935-1975*. Chicago, IL: American Library Association.
- Naroll, R. S., & Bertalanffy, L. V. (1956). The principle of allometry in biology and the social sciences. *General Systems Yearbook*, 1, 76-89.
- Nielson, B. (1981). Technological change and professional identity. In L. C. Smith (Ed.), *New information technologies - new opportunities* (Papers presented at the 1981 Clinic on Library Applications of Data Processing, Urbana-Champaign, 26-29 April 1981, pp. 101-113). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.
- Richardson, L. F. (1961). The problem of contiguity; an appendix to the statistics of deadly quarrels. *General Systems Yearbook*, 6, 139-187.
- Siegel, M. A. (Ed.). (1991). *Design and evaluation of computer/human interfaces: Issues for librarians and information scientists* (Papers presented at the 1988 Clinic on Library Applications of Data Processing, Urbana-Champaign, 5-7 April 1988). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois.
- Smith, L. C., & Dalrymple, P. W. (Eds.). (1992). *Designing information: New roles for librarians* (Papers presented at the 1992 Clinic on Library Applications of Data Processing, Urbana-Champaign, 5-7 April 1992). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois.
- Thompson, D. W. (1977). *On growth and form* (Abridged edition). London, England: Cambridge University Press.
- Weibel, S. (1992). Automated cataloging: Implications for libraries and patrons. In F. W. Lancaster & L. C. Smith (Eds.), *Artificial intelligence and expert systems: Will they change the library?* (Papers presented at the 1990 Clinic on Library Applications of Data Processing, Urbana-Champaign, 25-27 March 1990, pp. 67-80). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois.
- Woldenberg, M. J. (1971). *Allometric growth in social sciences* (Harvard Papers in Theoretical Geography, Technical Report No. 6). Cambridge, MA: Harvard University Graduate School of Design.
- Woldenberg, M. J. (1973). An allometric analysis of urban land use in the United States. *Ekistics*, 36(215), 282-290.

Geographic Information Retrieval and Spatial Browsing

Digital library (DL) projects are beginning to create very large-scale repositories of digital information on a wide range of topics. As with traditional print libraries, this information can be indexed and retrieved in a variety of ways, ranging from purely descriptive cataloging of items in the database and topical analysis of content, to more specialized methods of classification and description that exploit the characteristics of digital information. This paper will examine the problems and prospects of a class of retrieval and indexing methods that are particularly suited to digital library materials with geographic content or associations. The characteristics of spatial queries and some of the problems of uncertainty and approximation in spatial and geographic information retrieval are discussed. Requirements and a methodology for automatic indexing and georeferencing of text documents are then examined. A "state-of-the-art" examination of network access to georeferenced information is provided, and a specification language and tool for development of graphical interfaces to support geographic information retrieval and spatial browsing is also described. In conclusion, general issues and characteristics of georeferenced multimedia information systems are discussed.

INTRODUCTION

Many users of digital and print-based libraries have needs for information that are best approached from a geographical perspective. These users include scientists with research interests that range from the global changes brought about by the Greenhouse Effect and ozone depletion, global climate modeling and ocean dynamics, to the ecological characteristics of a region. They also include historians who require information on specific areas (at particular times), grammar school students working on class projects about particular cities and countries, and developers and city planners who must develop environmental impact reports for a given site.

Digital libraries have multiplied the types of georeferenced information sources (i.e., information with associated geographic coordinates) available beyond the traditional print and paper forms (maps, geographically indexed books, etc.) to include remote sensing data and images from satellites and aircraft, databases of measurements (e.g., temperature,

wind speed, salinity, snow depth, etc.) from specific geographic locations, and complex vector information such as topographic maps. They also store large amounts of digitized text and photographs from a variety of sources in a variety of formats and include other "multimedia" information such as digitized sounds and video in the DL database. Storage and access to specific items in a DL database is generally fast and efficient. However, access to the contents, and particularly to information relevant to the particular topical or subject needs of the user of such a database, is another matter altogether. Users of digital libraries need to be able to *search* for specific known items in the database and to *retrieve* relevant unknown items based on various criteria. This searching and retrieval has to be done efficiently and effectively, even when the scale of the database reaches the multiterabyte range (as is expected in the not-too-distant future). This implies that digital library objects must be *indexed* so that users can retrieve them by content. Effective user interfaces also must be designed so that the users can both *search* for items based on particular characteristics and *browse* the digital library for desired information.

This paper will examine the notion of Geographic Information Retrieval (GIR) in the context of digital libraries; in particular, it will focus on a particular class of indexing and retrieval methods appropriate to digital library materials with geographic content or associations. Geographic indexing and access has long been recognized as problematic in libraries (Holmes, 1990). It has largely relied on verbal designations of places, commonly depending on the Library of Congress Subject Headings and Name Authorities as a source (Brinker, 1962; Larsgaard, 1987). Graphical methods of access (such as the use of index map sheets in cartographic collections) have been much more rare.

This paper will examine some of the characteristics of georeferenced information and how such information can be incorporated into digital libraries. The intent is to raise a number of issues in the design and use of digital libraries with regard to search and retrieval of their geographically oriented contents. Not all of the issues raised will have obvious solutions nor shall I attempt to solve the myriad problems with such systems. Subsequent sections will define and describe the characteristics of geographic information retrieval and spatial querying, examine the characteristics and advantages of spatial browsing as a method of presenting a variety of georeferenced information in a coherent framework, and examine indexing and access creation for georeferenced sources including a discussion of some automatic georeferencing methods developed at Berkeley for the analysis and georeferencing of text materials. Other sections will examine graphical interfaces to support geographic information retrieval and spatial browsing, and provide a description of a prototype toolkit designed to aid in developing user interfaces for geographi-

cal applications such as spatial search and browsing. Finally, the conclusion will examine some general issues and characteristics of georeferenced multimedia information systems.

GEOGRAPHIC INFORMATION RETRIEVAL AND SPATIAL QUERYING

Geographic Information Retrieval is concerned with providing access to georeferenced information sources. This phrase is intended to convey a specialization of the term "Information Retrieval (IR)." It includes all of the areas that have traditionally formed the core of IR research with an emphasis, or addition, of spatially- and geographically-oriented indexing and retrieval.

There is often a distinction in the literature between IR and "data retrieval" of the type associated with database management systems (DBMS). In practice, the distinction is one of degree rather than of kind. Figure 1 shows a spectrum of various attributes of information retrieval and data retrieval. We will examine each of the attributes depicted in the figure and attempt to see where GIR falls on the continuum.

In information retrieval, the underlying model of providing access to documents is probabilistic (I will use the term "document" to represent any item of potential interest in a collection or database regardless of the content—text, images, maps, video, etc.—or the form—paper or digital). It is concerned with such subjective and indeterminate issues as whether (and to what degree) a document satisfies a user's need for information—i.e., whether it is relevant for that user and request. Data retrieval, on the other hand, is deterministic with regard to retrieval operations. If a document fulfills the conditions specified in the user's query, then it is by definition "relevant." In Geographic Information Retrieval, we are concerned with both deterministic retrieval (such as finding all data sets that contain information on a particular coordinate) and probabilistic retrieval (such as finding all towns near a major river).

Indexing is required for both efficient access to large databases and to organize and limit the set of elements of a database that are accessible. Most information retrieval systems derive their index elements from the contents of the items to be indexed. The derivation may be simple extraction (such as extracting keywords from a text), inferential extraction (such as mapping from text word to thesaurus terms), or it may be intellectual analysis and assignment of index items (such as assigning subject headings to a document). In data retrieval, the element itself, in its entirety, is the indexing unit. Obviously this spectrum is not really smooth or continuous since both types of indexing may be present within the same system. In GIR, both of these extremes are blended. There may be

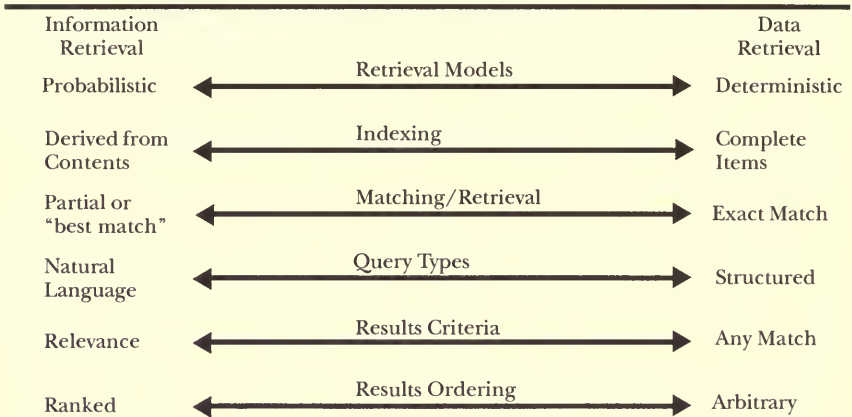


Figure 1. Information retrieval/data retrieval spectrum

intellectual indexing (such as assignment of bounding box coordinates to an aerial photograph) and inferential indexing (assignment of coordinates for places mentioned in a text).

In the actual retrieval of items from the database, the algorithms used for matching between the query and the index elements (or database contents) are based on the particular retrieval model. Information retrieval models lead to a class of retrieval algorithms that are probabilistic in nature and may involve the actual calculation of probabilities and use of statistical inference methods or may take an approach based on another model of the document space (such as Salton's [1989] vector space model). They attempt to find all of the potential (partial) matches between query and document and to rank them based on some measure of "goodness," so that the "best" matches receive the highest ranks. Data retrieval algorithms are deterministic and therefore demand an exact match between the query specification and the contents of the database. The Boolean logic used in processing the query languages of virtually all commercial database management systems, as well as in online catalogs and commercial information retrieval systems, is a deterministic algorithm. In GIR, both approximate partial matching and strict deterministic matching are of value in processing geographic and spatial queries (discussed further below).

The queries in information retrieval systems are commonly expressed as a natural language statement of the searcher's needs for information.

These queries are inherently imprecise and may be ambiguous. In data retrieval, the query is usually expressed in some sort of structured query language with precise syntactic and semantic characteristics. When the goal is to retrieve all items from a database that exactly match the specifications of the query, then there must be no ambiguity in the query statement as to what is wanted. The query types thus reflect the underlying models of the retrieval systems. In information retrieval, queries are taken as a "clue" as to what the searcher might consider to be a relevant item from the database, and retrieval is based on how well an item matches the clue. Typically, the results of a search are presented in a ranked order based on the degree of match between the query and the database item. In data retrieval, the query is taken as a precise specification of the desired items from the database, and retrieval is based on exact correspondence between the item and the query. Unless explicitly specified by the system or by the user as part of the query, there is no ranking or order imposed on the results of a data retrieval query.

Geographic Information Retrieval, as we define it here, is an applied research area that combines aspects of DBMS research, User Interface research, GIS research, and Information Retrieval research, and is concerned with indexing, searching, retrieving, and browsing of georeferenced information sources, and the design of systems to accomplish these tasks effectively and efficiently. In the next section, we will further examine the characteristics of geographic and spatial queries and where these fit on the continuum of Figure 1.

Geographic and Spatial Queries

The terms *geographic queries* and *spatial queries* imply querying a spatially indexed database based on relationships between particular items in that database within a particular coordinate system (or compatible coordinate systems). Spatial querying is the more general term. It can be defined as queries about the spatial relationships (intersection, containment, boundary, adjacency, proximity) of entities geometrically defined and located in space (De Floriani et al., 1993) without regard to the nature of the coordinate system. It could be argued that the Vector Space model of information retrieval (Salton, 1989) is a spatial querying system where the space and coordinates are defined by occurrence and frequency of term usage in a document collection. Geographic querying assumes that the space is delineated by the well-defined coordinate systems of the "real world." In the following discussion, the emphasis will be on geographic querying, although the underlying implementation might be a general purpose spatial database system rather than a geographic information system. As Frank (1991) has pointed out, there are many characteristics of geographic data that require special access methods and data

structures. We will not examine access methods here but will concentrate on a basic classification of types of spatial queries.

In general, geographical relationships in the coordinate systems imposed on the real world are geometric relationships. Within a geometric framework, where distance and direction can be measured on a continuous scale, many types of relationships between objects defined within that space can be determined using geometry. For example, given the geographic coordinates in latitude and longitude of Chicago (41°52'N 87°37'W) and New York (40°40'N 73°58'W), a fairly simple calculation can give the distance between the two cities (using the great circle method, the distance is $(1/2 \times 7915.6) \times (0.86838) \times (D\pi/180)$, where 7915.6 is the diameter of the Earth in miles, 0.86838 is the ratio of miles to nautical miles, and $\cos D = \sin \text{Latitude1} \times \sin \text{Latitude2} + \cos \text{Latitude1} \times \cos \text{Latitude2} \times \cos (\text{Longitude1} - \text{Longitude2})$, or about 651 nautical miles). Using the coordinates alone, it is simple to determine other relationships between the cities—e.g., Chicago is West and North of New York.

Spatial relationships may be both geometric and *topological* (spatially related but without measurable distance or absolute direction). Examples of topological relations include such properties as adjacency, connectivity, and containment. For example, whether some building is inside or outside the city limits of Chicago has to do with the building's relationship to an arbitrary boundary, but the distance or direction between the two is not an issue. Topological directions may have no particular relationship to any coordinate system that they might be imbedded in. "Left" and "Right" are valid directions only in relation to the observer's frame of reference and have no absolute relationship with "North" or "West."

Spatial and geographic queries combine both geometric and topological elements. Frew et al. (1995) suggest that there are two primary classes of requests from users: the "What's here?" query and the "Where's this?" query. The first type of query stems from a desire to discover what information is available about a particular location while the second stems from a desire to discover where certain phenomena occur. Within this simple classification of spatial and geographic queries, there are a number of different types of queries distinguishable by how the locations of interest are defined. The following discussion is based on the types of spatial queries defined by Laurini and Thompson (1992) and De Floriani et al. (1993).

Types of Spatial Queries

The types of spatial queries submitted by users to an information system such as a digital library may be arbitrarily complex in the types of information desired; the limitations on the areas, time periods, etc. covered; and many other conditions (spatial or not) that might be specified

in such a query. If we concentrate on only the spatial or geographic aspects of the query, there are a number of query types that can be distinguished based on the type of information provided by the user in the query. We will consider five types of spatial queries: (1) point-in-polygon queries; (2) region queries; (3) distance and buffer zone queries; (4) path queries; and (5) multimedia queries.

The last is actually a combination of multiple georeferenced sources in a single query. In the following discussion, we will examine each of these query types and their characteristics.

The first type of query is probably the most straightforward to process and describe. This is the *point-in-polygon* query (illustrated in Figure 2), which essentially asks the question “What do we have at this X,Y point in the current coordinate system?” The point-in-polygon query, in a digital library context, might ask which satellite images are available that show a particular spot or which documents describe the place indicated by the point. The query essentially asks for any georeferenced object or geographic data set that contains, surrounds, or refers to a particular spot on the surface of the earth. This is one of the more precise of all the spatial query types discussed here.

The next type of query is a *region* query (illustrated in Figure 3). A region query asks the question “What do we have in this region?” Instead of referring to a particular point in the coordinate space, a region query



Figure 2. Point in polygon

defines a polygon in that space and asks for information regarding anything that is contained in, adjacent to, or overlaps the polygonal area so defined. There are a number of potential variants or restrictions that might be applied—e.g., a user might ask “Which point encoded items lie within the region?” “What lines (borders, rivers, etc.) lie within or cross the region?” “Which areas (or regional data sets) overlap this region?” “Which areas (or regional data sets) lie entirely within this region?” or “Which areas share a border with this region?” Any combination of elements or containment criteria might be specified given the needs of the particular searcher. In addition, the specified query region can be any polygon ranging from regular shapes, such as rectangles or even circles (which would be the same as a Buffer Zone query on a point as discussed below), to irregular shapes, like the boundary of a city, or any arbitrary set of points defining a closed polygonal shape. The containment criteria need not be precise but may use “fuzzy” or probabilistic interpretations of such things as the maximum or minimum areas of overlap for an object to be considered included in the specified area, or the coverage areas for particular data sets that are candidates for retrieval (Brimicombe, 1993).

The next type of query is the *distance* and *buffer zone* query (illustrated in Figure 4). The distance and buffer zone query asks the question “What do we have within some fixed distance of this object (point, line, or polygon)?” Obviously, there are quite different processing steps involved if the object used as the basis for a buffer zone query is a point, a line, or a polygon. Examples include queries such as “What cities lie within forty miles of the border of Northern and Southern Ireland” (as shown in Figure 4)? Other buffer zone queries include: “What industrial



Figure 3. Region query



Figure 4. Buffer zone query

plants lie within two miles of this river?" "Which streams are within 100 yards of this highway?" "What mines are within five miles of this city?" etc. The buffer zone specified need not be exact—e.g., "What data sets describe the area near this point?" and inclusion can be considered a fuzzy or probabilistic function based on the location of the database objects. For such queries, a ranked list of database objects ordered by "nearness" to the point might be a better response than an arbitrary definition of a distance.

Path queries are a somewhat more specialized form of spatial query that require the presence of a network structure in the spatial or geographic data. Networks are simply sets of interconnected line segments representing such things as roads, oil or water pipelines, etc. A typical sort of path query involves finding the shortest route from one point in the network to another. For example, a path query might ask "What is the shortest route from San Francisco to Los Angeles" (as shown in Figure 5)? Note that path queries can become more complex (and uncertain) multimedia queries when criteria other than distance or direction are involved in the query. For example, to answer the question "What is the fastest route from San Francisco to Los Angeles?" more information, such as speed limits and traffic conditions on different routes of the network, is required to provide even an approximate answer.

Multimedia queries combine multiple georeferenced information sources in resolving a query. This may include multiple maps (or map layers, depending on the sort of system used to resolve the query); it may also include nonmap georeferenced information such as ownership



Figure 5. Path query

records for particular parcels of land. An example (illustrated in Figure 6) might be the query “What are the names of farmers affected by flooding in Monterey and Santa Cruz Counties?” Answering this query involves not only map information, such as county boundaries and river locations, but also cadastral information to show who owns particular parcels of land along the rivers in the areas affected by flooding. In this particular query, complex operations are likely to be required, such as combining aerial or satellite photographs or remote sensing data (showing the extent of the flooding) and map and cadastre information, often from different databases with different measurement scales and levels of detail.

The types of queries discussed above can be combined in a complex search. For example, “What streams and rivers flow through the county in which the town of Richmond (California) exists?” would require a point-in-polygon search of county information to locate the county containing the city and a region search to identify the streams and rivers that intersect the county area. Obviously, any GIR system should combine the text or concept-based retrieval associated with conventional information and database systems with the sorts of spatial queries discussed earlier. Any multimedia information system may include a wide variety of spatial and nonspatial information that may have a geographic association if not a precise location (see, for example, Griffiths, 1989). Walker, Newman, Medyckj-Scott, and Ruggles (1992) provide an interesting design for a system combining spatial, text, and concept-based retrieval.

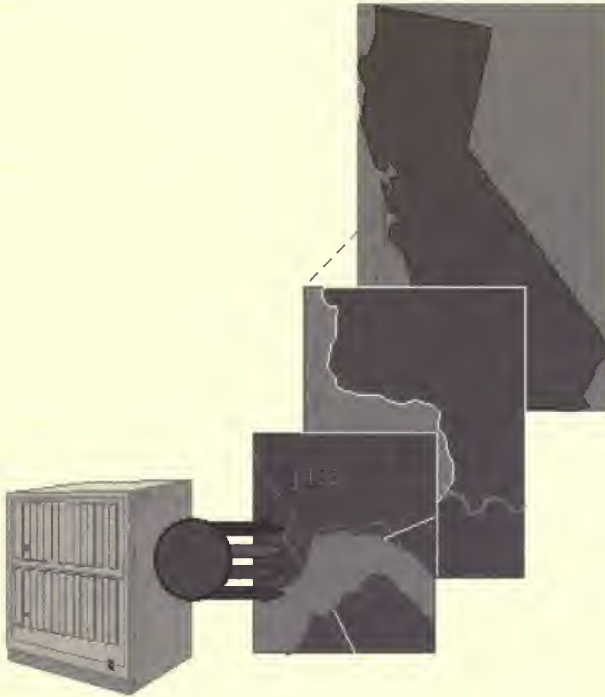


Figure 6. Multimedia query

SPATIAL BROWSING

Searching a geographically indexed database or digital library is an activity that assumes the searcher has an idea of what he or she wants and is able to specify that need in some form. Most of the queries used as examples in the above discussion reflect this. Another type of “searching” is much less directed, and, while it assumes that the users have some notion of the type of information desired, they may not be able to specify that information in a query language. What is needed in such cases is (in effect) the ability to navigate the database geographically without requiring explicit query formulation. This “spatial browsing” combines ad hoc spatial querying with interactive displays of digital maps to permit the user to explore the geographical dimension of information in a database or digital library.

Laurini and Thompson (1992) describe spatial browsing using the “hypermap” concept. In hypertext databases (the current best example being the World Wide Web), each document (or *node*) may contain many

links to other documents in a variety of media (text, images, video, sound clips, etc.), and the user may view any referenced document simply by selecting the representation of the link in the current document. In a hypermap, the links are represented by an *icon* or *footprint* (a polygon that outlines the area described by the object linked to the footprint), and selection brings up the document referenced by the link. For example, Figure 7 shows a sequence of maps that might be presented to someone browsing a digital library, going from a global view to Europe, then to the United Kingdom, then to Ireland, and finally to a particular icon on the Ireland map representing a book about the region.

There are a number of advantages to spatial browsing systems and the hypermap concept as a user interface “metaphor.” These systems are often very intuitive and comprehensible (assuming that the user has some notion of geography) and can provide for both searching and browsing by direct interaction as opposed to specification of names or coordinates. In most cases, for the purposes of browsing and search specification, the digital map displayed to the user need not be highly detailed, nor does it require the accuracy of a full GIS.

There are also a number of potential problems or requirements for such systems. One problem is that of clutter in the display. If the icons or footprints representing all of the documents in a large database associ-



Figure 7. Spatial browsing using the HyperMap

ated with any geographic area visible on the digital map are shown simultaneously, the map may disappear entirely beneath a heap of icons. This sort of clutter can be addressed in several ways, some of which we will discuss further below in describing the geographic browser toolkit. Another obvious requirement for spatial browsing is that there must be coordinate-based geographical indexing of the database. In the following section, methods of automatic indexing and automatic georeferencing of text documents will be examined. We will return to the notion of spatial browsing and examine some examples in subsequent sections.

GEOGRAPHIC AND SPATIAL INDEXING

Not surprisingly, one of the major sources of information in digital libraries is text in a variety of forms and from a variety of sources. These text items might include full-text documents—such as journal or encyclopaedia articles, books, technical reports—and more specialized documents—such as Environmental Impact Reports (EIRs), laws, and legislation. Many of these text documents describe, discuss, or refer to particular places or regions. Geographic location is often an important, or even the primary, criteria when searching for information from the digital library.

In traditional library cataloging practice, geographic references have been a common form of access point assigned to documents (primarily books and maps), but assignment was based on the cataloger's notion of whether geographic identification was deemed important for access to the document. Although it might be possible, in principle, to have catalogers evaluate each item that is entered into a digital library for geographic references in its content, such detailed cataloging would be prohibitively expensive. One goal of many digital library projects is to automate as much of the indexing and cataloging of documents as possible. An important component of such automatic indexing is to develop methods that can perform automatic *georeferencing* of text documents. By automatic georeferencing, I mean to automatically index and retrieve a document according to the geographic locations discussed, displayed, or otherwise associated with its content.

In most existing full-text and bibliographic information retrieval systems, searches with a geographical component, such as the point-in-polygon region or multimedia query ("locate any documents whose contents are about location XY"), are not supported directly by indexing, query, or user interface functions. Instead, these searches rely on indexing and query specification of place names either supplied by catalogers or extracted from the text itself, essentially as a side-effect of keyword indexing. Even in cases where a document is meticulously manually indexed,

geographic index terms consisting of text strings (such as LCSH and LC name authorities) have several well-documented problems with ambiguity, synonymy, and with name changes over time (Griffiths, 1989; Holmes, 1990). Specifically, the major problems are:

- Names are not unique: San Jose is a common city name throughout Central and South America as well as in California. Without additional qualifications, many place names are ambiguous.
- The places referred to change in size, shape, and in names over time: political changes in the world move much faster than geological changes, and borders, country and region names, even the existence of political entities, may change at any time.
- Spelling variations: Local names for a region may differ from common English forms, and there may be variations in the spelling of a name over time (Peking, Beijing).
- Some place names in texts are simply temporary conventions: in some scientific studies, as well as in some historical contexts, particular names may be created by scholars to describe an area or region (study areas, battlefields, etc.) that are not part of the conventional political names of a region but which may be very precisely defined for the purposes of the study.

Instead of, or in addition to, using place names to describe locations referred to in documents, digital libraries are using the geographic coordinates of places to provide better access to documents dealing with those locations. Geographic coordinates have several advantages over names:

- They are persistent regardless of name, political boundary, or other changes. A geographic location specified by coordinates is not dependent on the vagaries of politics, warfare, or synonymy.
- They can be simply connected to spatial browsing interfaces and GIS data. As discussed in subsequent sections, coordinate-based locations, or representation of documents associated with those locations, can be displayed and overlaid on digital maps.
- They provide a consistent framework for GIR applications and spatial queries: Having geographic coordinates for an object (whether specified as a point or as a polygonal region) as index entries permit precise or approximate spatial querying of the database using all the types of spatial searching discussed earlier.

The challenge is to provide reliable automatic indexing for geographic locations based on the names that occur in a text. Toward this end, the GIPSY system was developed at U.C. Berkeley by Ph.D. students Woodruff and Plaunt (1994a, 1994b).

GIPSY: Automatic Georeferencing of Text

GIPSY, The Georeferenced Information Processing System, was developed as a new model of automatic geographic indexing for text documents. In the GIPSY model, words and phrases containing geographic place names or geographic characteristics are extracted from documents and used to provide evidence for probabilistic functions incorporating elementary spatial reasoning and statistical methods to approximate the coordinates of the location being referenced in the text. The actual "index terms" assigned to a document are a set of coordinate polygons that describe an area on the Earth's surface in a standard geographical projection system. The GIPSY method for automatic georeferencing is described in detail by Woodruff and Plaunt (1994a) and will only be summarized here.

GIPSY uses a three-step algorithm which relies on a thesaurus or gazetteer containing place names and the names of other geographically significant objects (rivers, lakes, bioregions, animal and plant habitats, land use types, etc.).

Step 1: Identifying geographic place names and phrases. This step attempts to locate all relevant content-bearing geographic words and phrases in a text. This involves parsing the text using a parser that "understands" how to identify geographic terminology of two types:

1. Terms which exactly or closely match objects or attributes in the geographic thesaurus. This step requires a large gazetteer of geographic names and terms along with their geographic coordinates. Terms added to this thesaurus include generic terms for geological features, climate, land use, animal and plant species, and size.
2. Lexical constructs which contain spatial or topological information, such as "adjacent to the lake," "south of the river," "between the river and the highway," etc.

To implement this, a list of the most commonly occurring constructs must be created and integrated into a thesaurus/gazetteer.

Step 2: Locating pertinent data. The output of the first step is a set of extracted terms and phrases. In this second step, these terms and phrases are processed by a function which retrieves geographic coordinate data related to them. This step uses spatial data sets that provide information such as the names, sizes, and location of cities, states, etc.; names and locations of endangered species; names, locations, and bioregional characteristics of different climatic regions; etc. The system attempts to identify the spatial locations (a set of one or more geographic coordinates)

which most closely match the geographic terms extracted in the first stage. In some cases, where geographic modifiers are used, the area of coverage is modified to take into account the usage in the text. For example, the phrase "south of Lake Tahoe" might be mapped to the area south of Lake Tahoe and cover approximately the same volume. Since there is also geopositional data for land use (cities, schools, industrial areas, etc.) and habitats (wetlands, rivers, forests, indigenous species, etc.) available, extracted keywords and phrases for these types of data are also recognized in step one and locational information extracted in this step. The thesaurus entries for these data should incorporate several other types of information such as synonymy (e.g., Latin and common names of species) and membership (e.g., wetlands contain cattails but geopositional data on cattails may not exist, so we must use their mention as weak evidence of a discussion of wetlands and use these data instead).

In the prototype implementation of GIPSY, two primary data sets were adopted to construct the thesaurus and provide geographic locations. The first was a subset of the U.S. Geological Survey's Geographic Names Information System (GNIS) (USGS, 1985). The information extracted from the GNIS database contains latitude/longitude point coordinates associated with over 60,000 geographic place names in California. Data for land use and habitat data were derived from the U.S. Geological Survey's Geographic Information Retrieval and Analysis System (GIRAS) (Anderson et al., 1976).

The names and terms derived from the text may be associated with more than one location, so every identified name, phrase, or region description is associated with all of the coordinate points or polygons that *might potentially* be the place mentioned in the text. A probabilistic weight is assigned to each of these coordinate sets based on statistical information, such as the frequency of use of its associated term or phrase in the text being indexed and in the thesaurus. Many relevant terms do not exactly match place names, geographic features, or land use types in the database. Therefore, to accommodate these inexact associations between the text and the coordinate databases, the thesaurus was extended to include both manually inserted terms and by extraction of generic term relationships from the WordNet thesaurus (Miller et al., 1990) including synonyms, hyponyms, hypernyms, meronyms, holonyms, and evidonyms.

Step 3: Overlaying polygons to estimate approximate locations. Having identified many places associated with the terms extracted from the text and their variants, the next step is to attempt to infer the most likely geographic location(s) for the areas discussed in the text. Each geographic phrase, the probabilistic weight, and the coordinates derived in the preceding step can be represented as a three-dimensional "extruded" poly-

gon with its base in the plane of the x,z axes and which extends upward on the y axis a distance proportional to its weight (see Figure 8a). As new polygons are added, three cases may arise:

1. If the base of a polygon being added does not intersect with the base of any other polygons, it is simply laid on the base map beginning at $y = 0$ (see Figure 8b).
2. If the polygon being added is completely contained within a polygon which already exists on the skyline, it is laid on top of that extruded polygon—i.e., its base is positioned in a higher y plane (see Figure 8c).
3. If the polygon being added intersects, but is not wholly contained by, one or more polygons, the polygon being added is split and the intersecting portion is laid on top of the existing polygon and the nonintersecting portion is laid at a lower level. To minimize fragmentation in this case, polygons are sorted by size prior to being positioned in the “skyline” created by overlaying the polygons (see Figure 8d).

In effect, the polygons are “summed” by weight to form a geopositional “skyline” whose peaks approximate the geographical locations being referenced in the text. The geographic coordinates to assign to the text segment being indexed are determined by choosing a threshold of “elevation” z in the skyline, taking the x,z plane at z , and using the polygons at that “elevation.” Raising the threshold “elevation” tends to increase the accuracy of the retrieval while lowering it tends to include other similar regions (or regions described in the same way and a region discussed in a given text).

To show the results of this process in the GIPSY prototype, consider the following text from a publication of the California Department of Water Resources:

The proposed project is the construction of a new State Water Project (SWP) facility, the Coastal Branch, Phase II, by the Department of Water Resources (DWR) and a local distribution facility, the Mission Hills Extension, by water purveyors of northern Santa Barbara County. This proposed buried pipeline would deliver 25,000 acre-feet per year (AF/YR) of SWP water to San Luis Obispo County Flood Control and Water Conservation District (SLOCFCWCD) and 27,723 AF/YR to Santa Barbara County Flood Control and Water Conservation District (SBCFCWCD)....This extension would serve the South Coast and Upper Santa Ynez Valley. DWR and the Santa Barbara Water Purveyors Agency are jointly producing an EIR for the Santa Ynez Extension. The Santa Ynez Extension Draft EIR is scheduled for release in spring 1991.

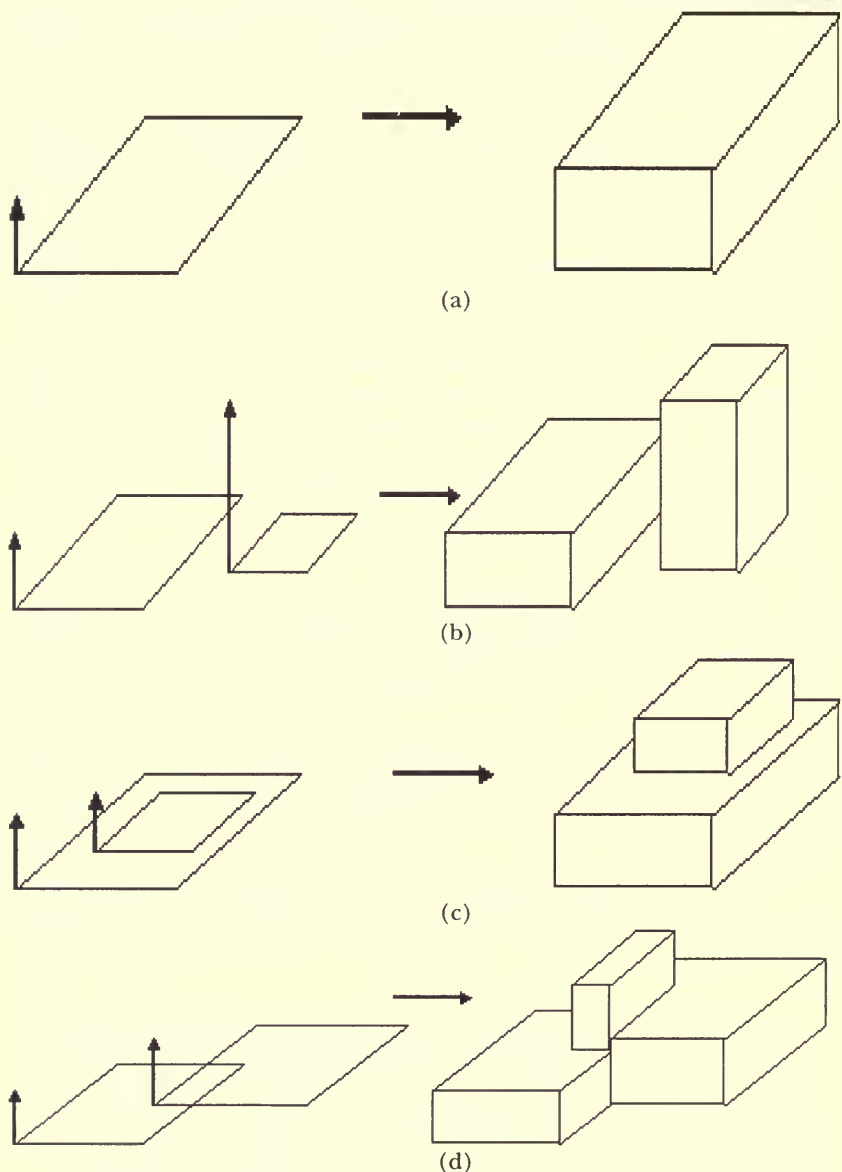


Figure 8a. The “weight” of a polygon, indicated by the vertical arrow, is interpreted as thickness or “elevation”.

Figure 8b. Two adjacent polygons do not affect each other; each is merely assigned its appropriate elevation.

Figure 8c. When one polygon subsumes another, their elevations in the area of overlap are summed.

Figure 8d: When two polygons intersect, their elevations are summed in the area of overlap.

Figure 9 contains a gridded representation of the state of California, which is elevated to distinguish it from the base of the grid. The northern part of the state is on the left-hand side of the image. The towers rising over the state's shape represent polygons in the skyline generated by GIPSY's interpretation of the text. The largest towers occur in the area referred to by the text, primarily centered on Santa Barbara County, San Luis Obispo, and the Santa Ynez Valley area.

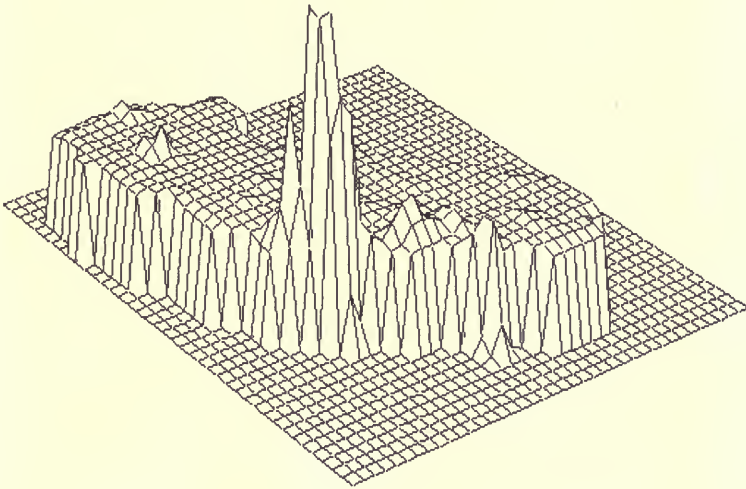


Figure 9. Results of polygon stacking from the DWR text

The surface plots generated in this fashion can also be used for browsing and retrieval. For example, the two-dimensional base of a polygon with a thickness above a certain threshold can be assigned as a coordinate index to a document. These two-dimensional polygons might then be displayed as icons or "footprints" on a map browser such as those discussed later. In addition, a natural language query describing an area of interest could be processed by the GIPSY system and candidate coordinate sets could be generated and ranked according to their weights and then used to retrieve georeferenced information located in those areas.

Ongoing research and development of the GIPSY system is being conducted at Berkeley in conjunction with the NSF/NASA/ARPA Digital Library Initiative project (Wilensky et al., 1994). We plan to use GIPSY as part of the automatic indexing mechanism for all texts stored in the digital library database.

EXAMPLES OF GIR ON THE INTERNET

Indexing texts and other georeferenced objects (such as photographs, videos, remote sensing data sets, etc.) by coordinates permits the use of the hypermap concept as discussed above. Use of coordinates also provides the capability of using a graphical interface to the information in the database or digital library where representations of the objects are displayed as icons or footprints on a digital map. As suggested earlier, a map-based graphical interface has a number of advantages over one which uses text terms or which simply uses direct specification of coordinates. It has been suggested that different cognitive structures are used by people in dealing with graphical and spatial information than those used for verbal information (Jones & Dumais, 1986), and that spatial queries cannot be adequately expressed by verbal queries (Furnas, 1991). Geographical queries are inherently spatial, and a map-based geographical interface tends to be intuitive and comprehensible to anyone who is somewhat familiar with maps and geography. Morris (1988) suggests that when users are given a choice between menu (text-based) and map-based graphical interfaces to a geographic database, they prefer the maps. A graphical interface, including digital maps and hypermaps, also permits dense presentation of information (McCann et al., 1988; DiBiase et al., 1993).

In this section, we will examine some examples of GIR and spatial browsing that are currently available via the World Wide Web (WWW). This "state-of-the-art" survey is by no means a comprehensive review of such systems but provides some examples to illustrate the concepts of GIR and spatial browsing as they have been implemented in current systems. The URL addresses of the systems discussed are indicated in square braces.

Figure 10 shows a hypermap from the UC Berkeley Environmental Digital Library prototype [<http://elib.cs.berkeley.edu>] (Wilensky et al., 1994). The base map shows the San Francisco Bay Area, and overlaid on the base map are three boxes showing the footprints of related hypermaps. Clicking a mouse button over the leftmost box brings up the more detailed map shown in Figure 11. Each of the labeled boxes overlaid on Figure 11 represents an aerial photograph (each image is actually a composite of digital remote sensing data acquired from an airborne platform) taken in order along a particular flightline. Clicking on the box marked 1-4 retrieves the image for the area centered under the box as shown in

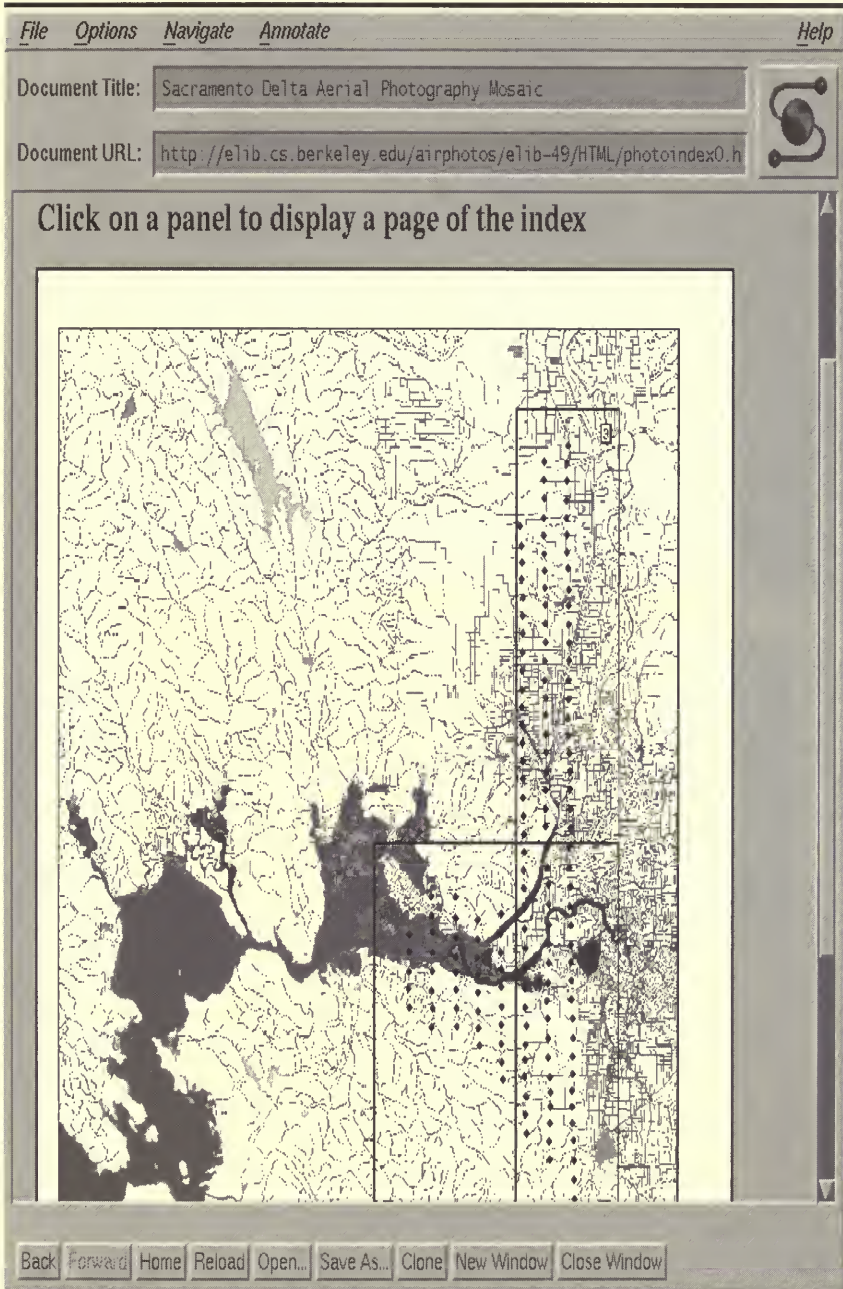


Figure 10. Hypermap from the UC Berkeley Digital Library Prototype



Figure 11. Detail map from the UC Berkeley Environmental Digital Library Prototype

Figure 12. The arrows above the image in Figure 12 permit the user to “navigate” up and down and through the images in a particular flightline, and left and right to images in parallel flightlines. In this prototype browser, the database is limited, and the maps are static entities. In the interface planned for the digital library, maps that can be scrolled and that provide the ability to “zoom” in and out will be used. A much more complex example of this type of hypermap access to information resources (jumping from static map to static map and eventually to data of interest) is provided by the “Virtual Tourist” system of the WWW that provides access to WWW servers by geographic location [<http://wings.buffalo.edu/world/>].

An example of the ability to scroll about and zoom in and out in a dynamic digital map is provided by the prototype version of the U.S. Bureau of the Census’ TIGER Mapping Service (TIGER stands for the Topologically Integrated Geographic Encoding and Referencing system, which was developed to support the activities of the Census Bureau starting with the 1990 census). Figure 13 shows a page from the TIGER Mapping Service prototype [<http://tiger.census.gov/>] that is centered on the New York City area. Using the four “arrows” surrounding the map, the user can move the view (pan or scroll) left or right, up or down.

The “zoom in” and “zoom out” buttons provide a reduced or increased scale of the map to be displayed. Figure 14 shows the results after “zooming in” on the area of Central Park. In Figure 14, much more detail (such as individual streets) is shown, and the legend has been changed to reflect the types of information that might be visible at the current scale. As noted on the WWW pages shown in Figures 13 and 14, the map that is shown is generated “on-the-fly” from the underlying TIGER digital mapping database. This means that instead of a static map (such as that in the digital library prototype discussed above), the maps in the TIGER Mapping Service are created on demand in response to a user’s mouse click on one of the scrolling or zooming buttons.

Figures 15 and 16 show an example of a full-scale geographic information system being made available via the WWW. The GRASSLinks system [<http://www.regis.berkeley.edu/grasslinks/>] was developed at UC Berkeley by Susan Huse (1995) as part of her Ph.D. work. It uses the GRASS (Geographic Resources Analysis Support System) GIS developed by the U.S. Army Corps of Engineers with an interface tailored to simple access over the WWW. Figure 15 shows part of the page that permits the user to specify the contents of the map desired, and Figure 16 shows the map generated by GRASSLinks (in this case a fairly simple map showing the counties of the San Francisco Bay area). GRASSLinks, like the TIGER system, allows the user to zoom in or out on particular areas, but it adds the additional capability to query the underlying geographic data sets



Figure 12. Aerial photograph from UC Berkeley Environmental Digital Library

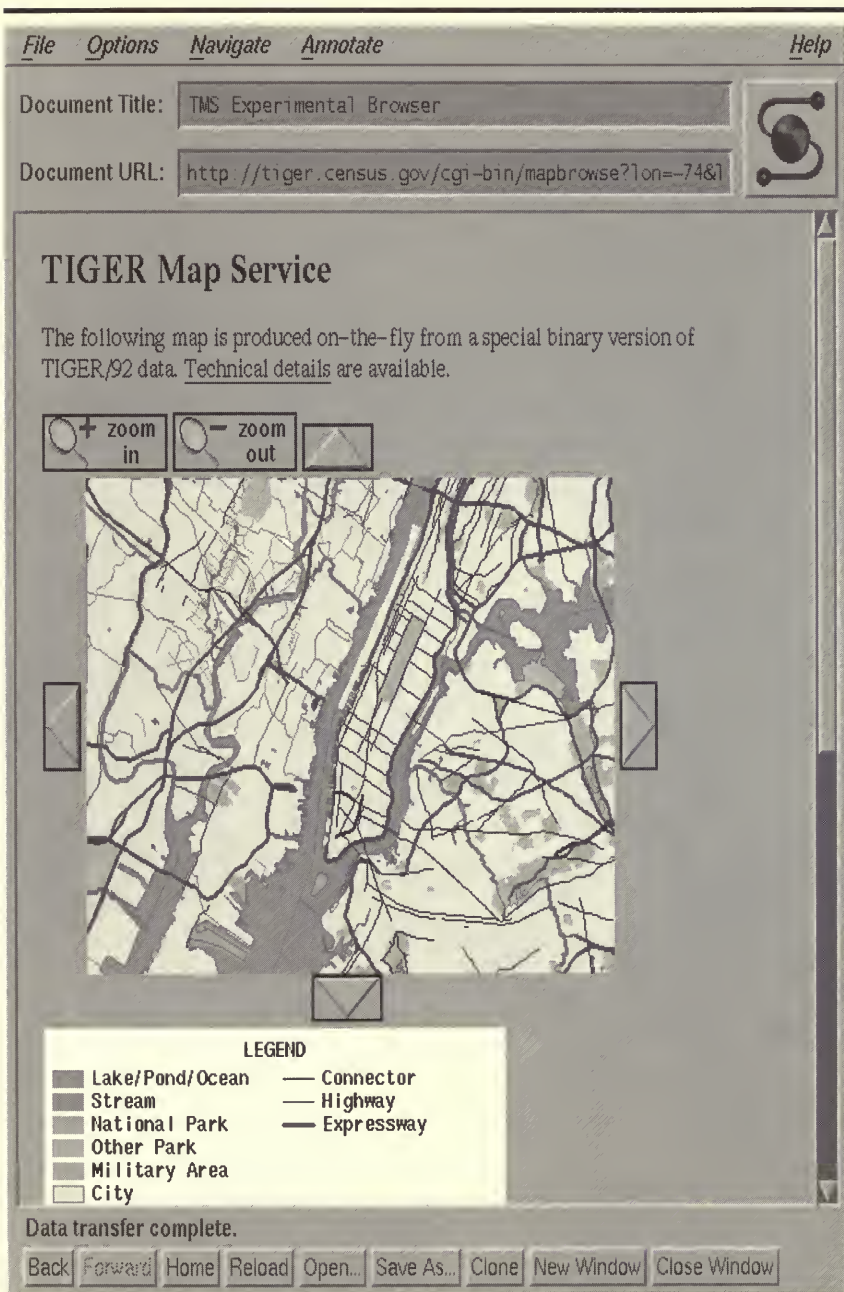


Figure 13. US Census Experimental TIGER Mapping System—New York Area

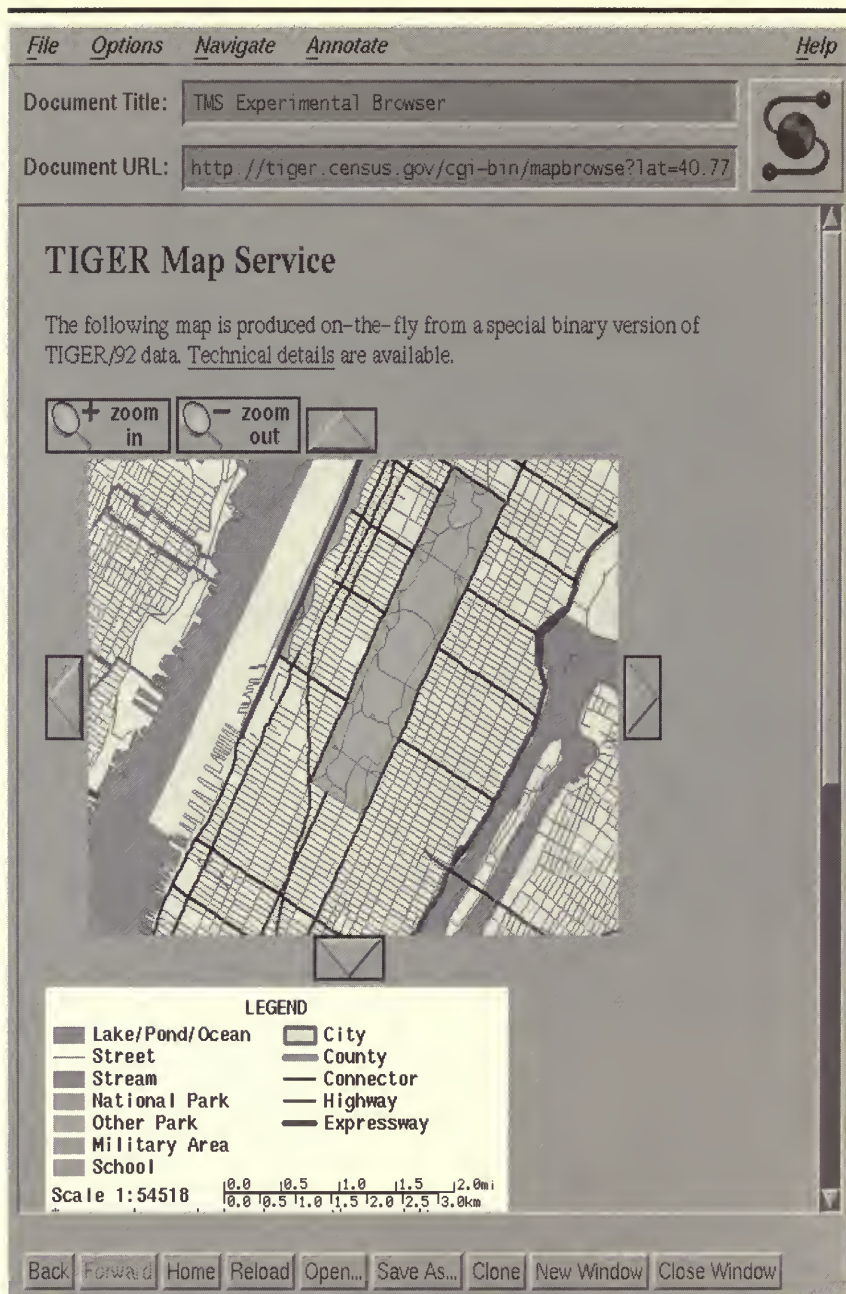


Figure 14. US Census Experimental TIGER Mapping System—New York area zoomed

FileOptionsNavigateAnnotateHelp

Document Title: Display Request Form

Document URL: http://www.regis.berkeley.edu/grasslinks/display.ht

Vector map display color: blue

c) Overlay sites map:

none
mines

Sites map display color: red

2) Select a region and resolution to be used.

(NOTE: run time varies with the resolution)

a) Region:

set to area of selected maps
12 County Bay and Delta Area
9 Bay Counties
3 Delta Counties
Sacramento San Joaquin Delta

b) Resolution: medium

3) Create the display image:

Display Maps

or return to the default values and start again: Reset

BackForwardHomeReloadOpen...Save As...CloneNew WindowClose Window

Figure 15. GRASSLinks specification page

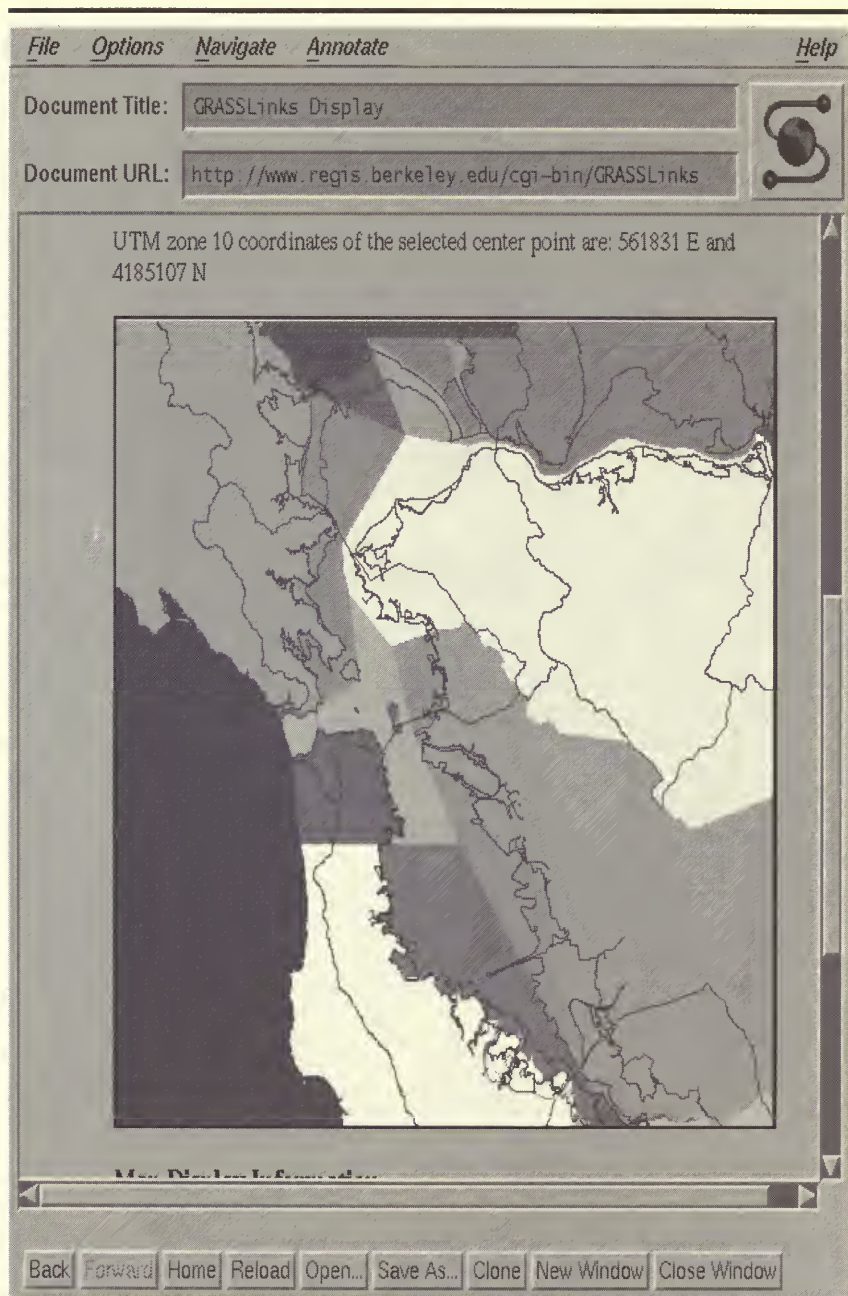


Figure 16. GRASSLinks map display

and display values (using a point-in-polygon search to identify the characteristics of an area surrounding a user-selected point), as well as to overlay multiple types of georeferenced data (ranging from data from the National Wetlands Inventory to geological maps of the San Francisco Bay area). A similar WWW interface to a somewhat less sophisticated GIS was developed for Canadian National Atlas data [<http://ellesmere.ccm.emr.ca/naismap/naismap.html>].

Not all GIR systems on the internet support graphical searching (at least at the present time). For example, the U.S. Geological Survey node of the National Geospatial Data Clearinghouse [<http://h2o.er.usgs.gov/nsdi/>] (part of the National Spatial Data Infrastructure [NSDI] effort by the federal government) has a GIR system for identifying relevant data sets that permits users to specify their requirement by both keywords and by specifying the latitude and longitude coordinates of a bounding rectangle for their area of interest (Nebert, 1995).

A wide variety of other geographically referenced (or related) information sources are available via the Internet. One of the best clearinghouses providing pointers to such information is the U.S. Geological Survey's "Geographic Information Referral Page" [<http://waisqvarsa.er.usgs.gov/wais/html/geog.html>].

Many of the network-based systems discussed earlier have some significant problems resulting from current technological limitations. One problem is that the time required for generating and transmitting maps such as those used in the TIGER Mapping System and the GRASSLinks system is much too long for convenient interactive use. Scrolling and zooming in the TIGER prototype may take as much as a minute to display the response to a single mouse click. In addition, the current version of the HTTP protocol and HTML "map widgets" used in the WWW for these systems have no way of transmitting continuous information from mouse movements or from multiple mouse selections of points on a map. This makes it impossible for current versions of WWW clients, such as Mosaic and Netscape, to have users interactively draw a bounding rectangle or polygon over the area they are interested in in order to submit a region query. Although support for such interaction is promised in the next version of HTTP, there are other problems with the HTTP protocol for use in information retrieval, including in GIR systems. The primary drawbacks are that the protocol is "stateless," meaning that each transaction between a client and server is treated as a completely distinct event, and the server keeps no information about clients from transaction to transaction. Some systems have finessed this limitation by adding state information to the query sent to the server, but these tend to be server-specific extensions to the protocol without standardization between servers.

There are other standard protocols, such as the Z39.50 information retrieval protocol (NISO, 1992), where the server does maintain a connection with the client, and which are better oriented to conventional information retrieval tasks. The National Geospatial Data Clearinghouse mentioned above is currently using the WAIS protocol based on an earlier version of Z39.50 and is moving to a full Z39.50 implementation with support for map browsing and bounding box selection (Nebert, 1995). Z39.50 is also being used for spatial indexes for the Government Information Locator Service [<http://www.usgs.gov/gils/>]. In the following section, we describe a toolkit for building geographical browsing interfaces that also provide Z39.50 access to remote servers of geographic and georeferenced information. Another possible solution is the Java programming language which permits Java-compatible WWW browsers to download and execute interactive programs (van Hoff, 1995).

GEOGRAPHIC BROWSING TOOLKIT

Most Geographic Information Systems were designed to support the work of cartographers, city and regional planners, and others with a need to *create* maps. As such, they tend to be "heavyweight" systems that often have a monolithic structure, proprietary data formats, and a tendency to be very complex to use. As such, these systems are ill-suited for applications where a user wishes to simply utilize spatial browsing to examine a variety of georeferenced information.

In this section, we will examine the features of a toolkit for constructing such browsing interfaces developed by the author that provides a simple solution to rapid development and presentation of geographical information retrieval systems. We will first examine the background on the toolkit and then describe one application using it and how a user might interact with that application. This is followed by a more detailed discussion of the toolkit contents.

In the Sequoia 2000 project, which was concerned with the design and development of very large-scale information systems for Earth scientists studying global change phenomena, a new browser paradigm was proposed to provide access to database information using a spatial paradigm (Chen et al., 1992). In this system, now called Tioga (Stonebraker et al., 1993), information is displayed topologically according to continuous characteristics which are attributes of the data. For example, documents may be displayed on a map according to their latitude and longitude. Documents might also be displayed according to the time at which they were generated and the time to which they refer as well as by more abstract functions such as the reading level of the document, the author's attitudes as expressed in the document, etc. A prototype of the geographi-

cal browsing component for Tioga was developed by the author and was included in the Lassen Geographic Browser (shown in Figures 17-19) as part of an interface for document retrieval that included both known item searching and probabilistic retrieval based on document contents (Larson, 1994). The current version of the geographic browser toolkit is a revised and refined version of the Lassen browser.

Using the Lassen browser, any georeferenced object in the database can be indicated by an icon on the map. The user employs the mouse to center the map on any location and to zoom in or out for more or less detailed maps, either by clicking on the map with the mouse or by moving the scroll bars below the map to set particular coordinates and zoom levels. With local processing handling all map drawing and updates, the Lassen Geographic Browser is fast enough to redraw the entire world map in less than a second. When zoomed in on a location, map updates are faster due to hidden line clipping. At reduced levels of detail, the browser can redraw the screen fast enough to provide an animated "real-time" zooming and scrolling effect. Icons can be placed at any coordinates on the map and for any range of "zoom" values. The icon will only appear when the coordinates are visible and the current "zoom" level is within the specified range. This helps to reduce clutter on the screen. Clutter is further reduced by placing references to multiple objects that deal with the same area into a pop-up menu. When an icon is selected by the user, the menu of objects georeferenced at the icon coordinates and detail level are displayed for selection.

In Figure 17, the initial global view of the Lassen browser is shown in a typical X Window system display window. By clicking with the left mouse button on the United States, the map was centered on that, clicking the middle mouse button zooms in one unit (and the right mouse button zooms out). The units used in zooming are logarithmic so that zooming from a global scale to the scale shown in Figure 18 requires only a single mouse click, but as the view "closes in," more clicks are needed to increase the zoom by the same amount. The convention adopted for displaying icons in this version of the browser is that the underlying objects represented by the icons should represent geographic coverage approximately equal to the area shown in the map window. In Figure 18, a single icon is shown as a small box in the center of the United States. When the icon is selected, a pop-up menu listing data objects associated with the whole area is visible including such things as satellite images of the entire region. Selecting an object from the menu invokes a viewer appropriate to the type of data (the Lassen prototype shown here included georeferenced texts, images in a variety of formats, and MPEG videos). Figure 19 shows a "closer" view centered on Southern California. In this window, the "place names" button was selected to show the names of cities



Figure 17. Lassen browser—Global view

in the visible area (the prototype uses a small gazetteer of international city names, which is not very accurate as to geographic location). When a user enters a place name at the bottom of the screen, the name is searched in the small international gazetteer and in a much more comprehensive and detailed U.S. gazetteer. When a name is found, the map is centered on the referenced location and zoomed to intermediate scale with the name displayed. The Lassen prototype browser shown in Figures 17-19 uses the geographic browser toolkit to display and manipulate the map information shown in the main window. The toolkit has been developed as an extension to the Tcl/Tk language (Ousterhout, 1994), developed to provide the ability to create interactive scripts for applica-



Figure 18. Lassen browser—Centered and zoomed on U.S.

tions. The toolkit shares all of Tcl/Tk's ability to simply construct X Window-based interfaces using simple commands. For example, to create the map shown in the main window only required one statement in a Tcl script:

```
map .mp.win1.mapwin -mapfile map.data -width 7i -height 3.5i
-detail 4 -aspect 100
```

This statement starts with the “verb” **map** that indicates a request for a new map “widget” (the term used in Tcl/Tk for all interface elements such as sliders, buttons, and text entry boxes). The verb is followed by



Figure 19. Lassen browser—Futher zoomed to Southern California with Place Names activated

the name to be used for the widget window created. Tcl/Tk uses an object-oriented style of object creation with all subsequent operations on an object treated as commands, where the “verb” is the name of the object involved, called “widget commands.” In the statement shown above, the widget name is followed by a number of options specifying some characteristics of the map. The **-mapfile** option indicates the file name of where the map data are located. The browser currently uses data in a compressed format derived from the World Data Bank II data set originally created by the Central Intelligence Agency and released to the public domain. Additional formats are planned for support of such things as TIGER map data. The next option in the command above (**-width** and

-height) provides the dimensions of the widget and may be specified in inches, centimeters, millimeters, printer's points, or screen pixel units. The **-detail** option indicates how detailed the map drawing should be, ranging from one (representing the highest detail available), to five (indicating that only about 1/5 of the detail will be displayed). The **-aspect** option is used to correct the appearance of maps when zoomed in at high and low latitudes (the map widget uses a cylindrical projection so there is a "spreading out" of features when drawing high and low latitudes). There are a number of additional options that can be specified in the map command. These include:

1. **Latitude** and **Longitude** for the centerpoint of the display, these default to 0°0'0"N and 0°0'0"W.
2. **Zoom** to specify the level of detail to be shown.
3. **MapBackground** to specify the color of the window background (default is black).
4. **Coastlines** to specify the color of continental coastlines (default is green).
5. **Countries** to specify the color used for national boundaries (default is red).
6. **States** to specify the color used for states in the U.S. and for international zones (default is yellow).
7. **Islands** to specify the color of island coastlines (default is light green).
8. **Lakes** to specify the color of lake shorelines (default is light blue).
9. **Rivers** to specify the color of river shorelines (default is blue).

Any of these options can be changed after the initial creation of a widget using the **map** command and by using the **configure** "widget command." For example, to change the color of the coastlines displayed from green to violet in the map widget created above (named `.mp.win1.mapwin`), the command:

```
.mp.win1.mapwin config -coastlines violet
```

could be issued in a script. In addition, multiple map widgets can be used in an interface. In Figure 20, for example, an interface with two map widgets is shown. The main map widget is the primary display, while the smaller widget is used to move the view of the main map widget to center on a point "clicked" in the small widget. Additional widget commands are used for a variety of operations, such as recentering the view of an existing map widget. These widget commands include:

1. **CenterPixel**—this tells the widget to redraw itself centered on the point indicated by *xy* pixel coordinates of the window supplied as an argument.

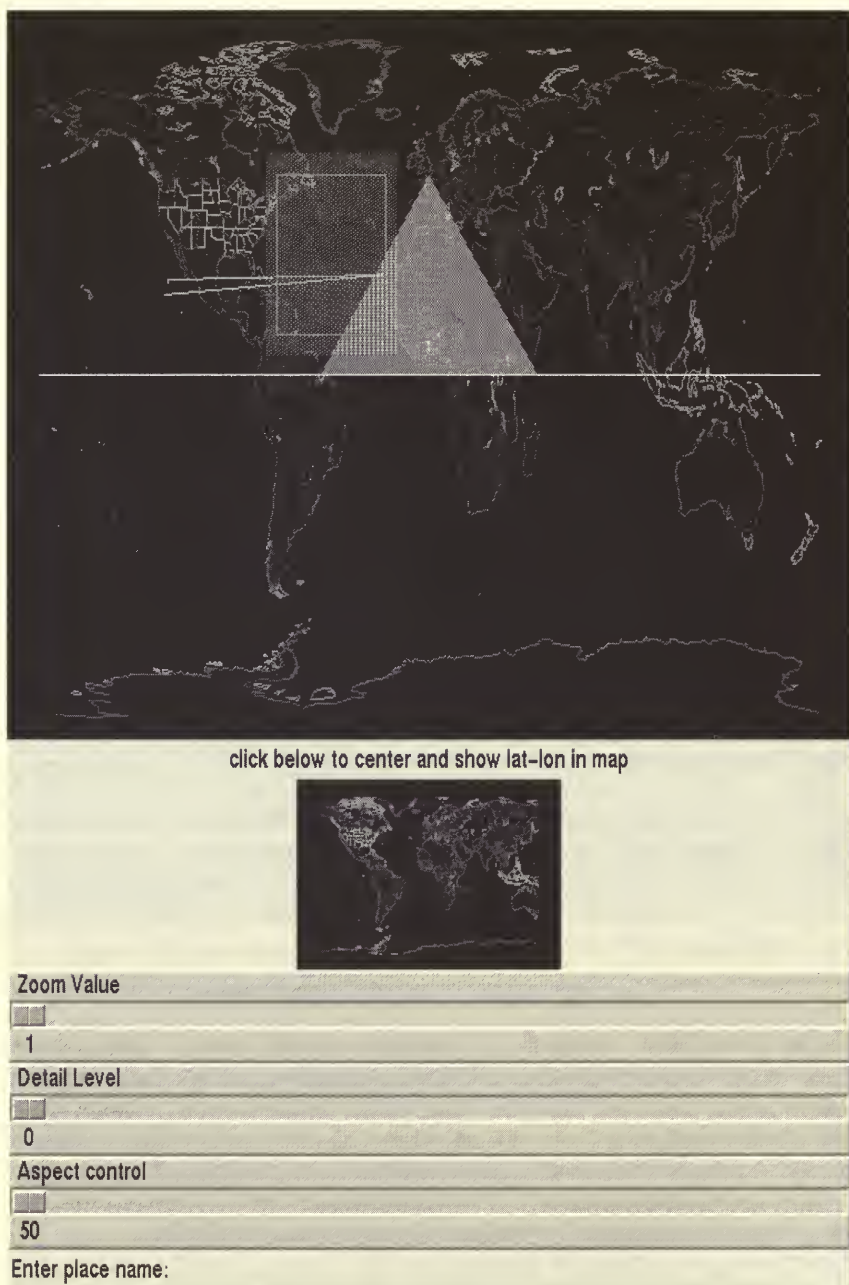


Figure 20. Map Widget example, with overlays

2. **Display** or **ReDisplay**—causes the widget to redraw itself immediately.
3. **ZoomMap**—Zooms in or out of the current map display to the “zoom value” supplied as an argument. This can be an absolute value or an increment to the current value (indicated by a plus sign) or a decrement to the current value (indicated by a minus sign).
4. **GetCenter**—this asks the widget to report the current latitude and longitude of the center of the map display.
5. **GetFrame**—this asks the widget to report the current latitudes and longitudes of the top-left and bottom-right corners of the map display.
6. **GetLatitude**—this asks the widget to report the current latitude of the center of the map display.
7. **GetLongitude**—this asks the widget to report the current longitude of the center of the map display.
8. **GetLatLon**—this asks the widget to report on the latitude and longitude of a point indicated by *xy* pixel coordinates of the window supplied as an argument.
9. **GetXY**—this asks the widget to report on the *xy* pixel coordinates of the window that represents the location specified by a latitude and longitude supplied as arguments.

In addition to these widget commands, there are three others: **Create**, **Destroy**, and **Rubberband**. **Create** creates a geometric object and places it in the map at the geographic coordinates specified in the create command; it also returns a unique identification number for the object. The **Destroy** command takes an object's identification number as an argument and removes the object from the display and from memory. The **Rubberband** command functions the same as a create command but permits dynamic changes to the coordinates of the object that are displayed. It is intended for interactive drawing of objects to show, for example, a rectangle that “stretches” over the base map following mouse movements until the mouse button is released. A **Rubberband** command with the single argument “**end**” finishes the interactive drawing of an object and automatically connects the final point of a polygon with its start point. Each object created remains attached to the geographic coordinates specified in the create command so that scrolling the display or zooming in or out will redraw the object appropriately. There are several types of objects supported in the current version of the map widget, including:

1. **point**—a single point indicated by latitude and longitude;
2. **line**—a straight line defined by the latitudes and longitudes of its end points;
3. **rectangle**—a rectangle defined by latitudes and longitudes of its top-left and bottom-right corners;

4. **filledrectangle**—a rectangle defined by latitudes and longitudes of its top-left and bottom-right corners and filled with a transparent color;
5. **polygon**—an arbitrary polygon defined by a set of latitudes and longitudes of each vertex and filled with transparent color; and
6. **polyline**—an arbitrary set of connected line segments defined by a set of latitudes and longitudes of each vertex and filled with transparent color.

Circles, ovals, and arcs are not supported in the current version, but support is planned for the future. The color of each object is specified in its **create** statement. These colors, like all colors in Tcl/Tk, can be specified using X color names or by specifying color values as 4-bit, 8-bit, 12-bit, or 16-bit hexadecimal values for the red, green, and blue components of the color. Most X Window servers have a rich variety of color names (e.g., BrickRed1, LightSkyBlue3, NavajoWhite1, SeaGreen2) including all common color names like red, green, black, and so on. For example, the following statements were used to create the objects shown on the main map in Figure 20:

```
.map create line white 0o0'0"N 179o0'0"W 0o0'0"N 179o59'59"E
.map create polygon white 50o0'0"N 0o0'0"W 0o0'0"N 50o0'0"W
0o0'0"N 50o0'0"E 50o0'0"N 0o0'0"W
.map create polygon blue 23o0'0"N 120o0'0"W 24o0'0"N 120o0'0"W
25o0'0"N 22o0'0"W 20o0'0"N 122o0'0"W 23o0'0"N 120o0'0"W
.map create polyline yellow 23o0'0"N 120o0'0"W 24o0'0"N 120o0'0"W
25o0'0"N 22o0'0"W 20o0'0"N 122o0'0"W 23o0'0"N 120o0'0"W
.map create point white 23o0'0"N 120o0'0"W
.map create rectangle red 50o0'0"N 70o0'0"W 10o0'0"N 20o0'0"W
.map create filledrectangle violet 55o0'0"N 75o0'0"W 5o0'0"N
15o0'0"W
```

Note that latitude and longitude are expressed in the conventional "degrees° minutes' seconds", direction (N, S, E, W) format, slightly adapted to permit entry on an ascii keyboard (lowercase *o* is substituted for the degree symbol). The toolkit also contains conversion routines to alter a variety of geographic coordinates into latitude and longitude.

Naturally, additional statements other than the one shown above are required to group and arrange the elements of the interface and to specify the characteristics and behavior of the other interactive elements on the screen, but Tcl/Tk handles a surprising amount of this work automatically. Any widget type, including maps, can have actions bound to X Window events such as mouse movement or mouse button presses, keystrokes, window movement, etc. For example, the Tcl statements:

```
bind .map <1> "%W centerpixel %x %y"
```

```
bind .map <2> "%W zoom +1.0 ; inczoom"
bind .map <3> "%W zoom -1.0 ; deczoom"
```

attach particular actions to mouse buttons 1, 2, and 3 when clicked on the map widget called **.map**. The current window name is substituted for %W by the bind command, and the current mouse *xy* location in window pixel units is substituted for %*x* and %*y* in the statements shown. Thus, the first line above sets up the map widget so that the view is centered on the point indicated by the current mouse position whenever mouse button one (the left) is clicked on the map widget. The next lines attach zooming in and out to the middle and right mouse buttons, respectively. Any arbitrary Tcl/Tk script can be included as the last element of the bind command. In the last line above, there are two Tcl statements; the first handles the zoom widget command and the second is an invocation of a script procedure named "deczoom" that handles updating the other widgets on the screen (such as the "zoom value" slider widget) to reflect the new zoom value. Any other Tcl/Tk widgets, such as buttons, text or graphical labels, menus, scales, and scroll bars can be placed anywhere on the map display (using the **GetXY** widget command to convert latitude and longitude to the appropriate pixel location on the map display). These other widgets can also have any arbitrary script attached to them.

Tcl/Tk and the geographic browser toolkit provide a convenient and flexible way to prototype and implement user interfaces for geographic information retrieval and spatial browsing applications. It is being used in an updated version of the Lassen geographic browser and is also being included in the Cheshire II next-generation online catalog system (Larson et al., 1995). Planned additions to the toolkit include more map data formats, and the ability to express coordinates in a larger variety of coordinate systems (e.g., Universal Transverse Mercator) and formats (e.g., latitude and longitude expressed as positive or negative decimal degrees instead of degrees, minutes, and seconds).

CONCLUSIONS

Digital libraries are multimedia information systems that will include digital "documents" in a variety of formats. In the Environmental Digital Library being constructed at Berkeley (Wilensky et al., 1994), we are building a digital collection that includes text documents such as Environmental Impact Reports, County General Plans, magazines and publications of the California State Resources Agency in both scanned page images, ASCII text form from Optical Character Recognition, and a number of other forms. The collection also includes scanned 35mm slides of places in

California from a collection of over 500,000 (with about 10,000 converted to digital form at the present time), and videos produced by state agencies. In addition, it also contains USGS topographic data, U.S. Census TIGER data, the contents of a large-scale GIS for the San Francisco Bay Area, and collection databases for classification and distribution of California flora and fauna. All of these data, in one way or another, have a geographic component to be considered in indexing and retrieval. We believe that geographic browsing will play a key role in providing effective access to this information, permitting scholars, scientists, government agency personnel, librarians, and students to quickly and easily find and retrieve the information they need for tasks ranging from research to environmental planning. To implement such a digital library and make it available to such a user population, a number of important research and development questions need to be examined. These include:

1. Is it possible to create a coherent content-based view of such a diverse collection of digital objects? Geographic Information Retrieval and Spatial Browsing, as discussed in this paper, can provide at least one coherent framework within which to coordinate many disparate resources in such a system.
2. How large can such a digital library be and still provide efficient and effective access to its contents? That is, how well do all the components of the digital library scale up in size from small prototypes to multiterabyte databases with hundreds of simultaneous users?
3. How can the contents and indexes of a digital library be distributed among many different computers over a wide area network and still provide coherent and comprehensive access to its contents? How will such distributed systems communicate and share information, and how will they communicate with other digital libraries and databases? How can such large amounts of data be transmitted most effectively and efficiently?
4. How will information be added to the digital library, what forms will these documents take, and how will content and characteristics of the documents be defined? Can these processes be automated and how effective will such procedures and techniques be in providing access to the collection? There is much work to be done on effective automatic indexing and classification for both full-text documents and the other sorts of digital objects included in the digital library.
5. What are the modes of interaction with the digital library? As suggested earlier, GIR and spatial browsing will play a large role. But more conventional information retrieval and DBMS querying techniques will also be required.

These open questions hint at the outlines of the research being conducted in support of the development of digital libraries. The Berkeley Digital Library Project and the Alexandria Digital Library Project at UC Santa Barbara (Frew et al., 1995) have adopted a geographical information retrieval approach as one of the primary interaction paradigms for digital library access. We are just beginning research into the problems, effectiveness, and usability of geographic information retrieval and spatial browsing, but I believe that a GIR and a spatial browsing paradigm will provide an elegant and comprehensible method of interaction for a wide variety of users and needs.

ACKNOWLEDGMENTS

Some of the research described here, including GIPSY and the Lassen prototypes, was sponsored by the University of California and Digital Equipment Corporation as part of the Sequoia 2000 project under DEC Research Grant #1243. Current work on GIPSY and the browser toolkit are being carried out under the auspices of the UC Berkeley Environmental Digital Library Project sponsored by NSF, NASA, and ARPA.

REFERENCES

- Anderson, J. R.; Hardy, E. E.; Roach, J. T.; & Witmer, R. E. (1976). *A land use and land cover classification system for use with remote sensor data* (Geological Survey Professional Paper #964). Washington, DC: USGS.
- Brimicombe, A. J. (1993). Combining positional and attribute uncertainty using fuzzy expectation in a GIS. In *GIS/LIS Proceedings* (pp. 72-81). Bethesda, MD : American Society for Photogrammetry and Remote Sensing, American Congress on Surveying and Mapping.
- Brinker, B. (1962). Geographic approach to materials in the Library of Congress subject headings. *Library Resources and Technical Services*, 6(1), 49-64.
- California Department of Water Resources. (1991). *State water project, Coastal Branch, phase II and Mission Hills Extension*. Sacramento, CA: California Department of Water Resources.
- Chen, J.; Larson, R. R.; & Stonebraker, M. (1992). Sequoia 2000 object browser. In *Digest of Papers. COMPCON Spring 1992. Thirty-Seventh IEEE Computer Society International Conference* (pp. 389-394). Los Alamitos, CA: Computer Society Press.
- De Floriani, L.; Marzano, P.; & Puppo, E. (1993). Spatial queries and data models. In A. U. Frank & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS* (#716 Lecture Notes in Computer Science). Berlin: Springer-Verlag.
- DiBiase, D.; Reeves, C.; MacEachren, A. M.; Krygler, J. B.; von Wyss, M.; Sloan, J. L.; & Detweller, M. C. (1993). A map interface for exploring multivariate paleoclimate data. In *Auto-Carto 11 Proceedings* (pp. 63-71). Bethesda, MD: The American Society for Photogrammetry and Remote Sensing.
- Egenhofer, M. J., & Frank, A. U. (1988). Designing object-oriented query languages for GIS: Human interface aspects. In *Proceedings: Third International Symposium on spatial data handling* (pp. 79-95). Columbus, OH: International Geographical Union, Commission on Geographical Data Sensing and Processing, Dept. of Geography, Ohio State University.
- Frank, A. U. (1991). Properties of geographic data: Requirements for spatial access methods. In O. Guenther & H. J. Schek (Eds.), *Advances in spatial databases* (Lecture Notes in Computer Science #525) (pp. 225-234). Berlin: Springer-Verlag.

- Frew, J.; Carver, L.; Fischer, C.; Goodchild, M.; Larsgaard, M. L.; Smith, T.; & Zheng, Q. (1995). The Alexandria Rapid Prototype: Building a digital library for spatial information. In *1995 ESRI User Conference Proceedings*. Redlands, CA: Environmental Systems Research Institute, 1995. <http://www.esri.com/resources/userconf/proc95/to300/p255.html>
- Furnas, G. W. (1991). New graphical reasoning models for understanding graphical interfaces. In S. P. Robertson, G. M. Olson, & J. S. Olson (Eds.), *Human factors in computing systems: Reaching through technology, CHI '91 Conference Proceedings* (pp.71-78). New York : ACM.
- Griffiths, A. (1989). SAGIS: A proposal for a Sardinian geographic information system and an assessment of alternative implementation strategies. *Journal of Information Science*, 15(4/5), 261-267.
- Holmes, D. O. (1990). Computers and geographic information access. *Meridian*, 4, 37-49.
- Huse, S. (1995). *GRASSLinks: A new model for spatial information access in environmental planning*. Unpublished doctoral dissertation, University of California, Berkeley.
- Jones, W. P., & Dumais, S. T. (1986). The spatial metaphor for user interfaces: Experimental tests of reference by location versus name. *ACM Transactions on Office Information Systems*, 4(1), 42-63.
- Larsgaard, M. L. (1987). *Map librarianship: An introduction* (2d ed.). Littleton, CO: Libraries Unlimited.
- Larson, R. R. (1994). Design and development of a network-based electronic library. In *Navigating the networks: Proceedings of the ASIS midyear meeting* (pp. 95-114). Medford, NJ: American Society for Information Science, Learned Information.
- Larson, R. R.; Moon, R.; McDonough, J.; Kuntz, L.; & O'Leary, P. (1995). Cheshire II: Design and evaluation of a next-generation online catalog system. *Proceedings of the American Society for Information Science Annual Meeting, 1995* (pp. 215-225). Medford, NJ: American Society for Information Science, Information Today.
- Laurini, R., & Thompson, D. (1992). *Fundamentals of spatial information systems*. San Diego, CA: Academic Press.
- McCann, C. M.; Taylor, M. M.; & Tuori, M. I. (1988). ISIS: The interactive spatial information system. *International Journal of Man-Machine Studies*, 28(2-3), 101-138.
- Miller, G. A.; Beckwith, R.; Fellbaum, C.; Gross, D.; & Miller, K. (1990). *Five papers on WordNet* (CSL Report 43). Princeton, NJ: Princeton University, Cognitive Science Laboratory.
- Morris, B. (1988). CARTO-NET: Graphic retrieval and management in an automated map library. *Bulletin of the Special Libraries Association, Geography and Map Division*, 152, 19-35.
- NISO - National Information Standards Organization. (1992). *Information retrieval service protocol for open systems interconnection* (ANSI Z39.50-1992). Bethesda, MD: NISO.
- Nebert, D. D. (1995). *Serving digital map information through the World-Wide-Web and Wide-Area Information Server Technology*. Reston, VA: U.S. Geological Survey. (<http://h2o.er.usgs.gov/public/WWW.paper.html>)
- Ousterhout, J. K. (1994). *Tcl and the Tk toolkit*. Reading, MA: Addison-Wesley.
- Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.
- Stonebraker, M.; Chen, J.; Nathan, N.; Paxson, C.; & Wu, J. (1993). Tioga: Providing data management support for scientific visualization applications. In R. Agrawal, S. Baker, & D. Bell (Eds.), *Proceedings of the 19th International Conference on Very Large Data Bases* (pp. 25-38). Palo Alto, CA: Morgan Kaufmann Publishers.
- United States Department of the Interior, U.S. Geological Survey. (1985). *Geographic Names Information System: Data users guide 6*. Reston, VA: U. S. Geological Survey.
- van Hoff, A. (1995). Java and Internet programming. *Dr. Dobbs's Journal*, 20(8), 56-61.
- Walker, D. R. F.; Newman, I. A.; Medyckyj-Scott, D. J.; & Ruggles, C. L. N. (1992). A system for identifying datasets for GIS users. *International Journal of Geographical Information Systems*, 6(6), 511-527

- Wilensky, R. et al. (1994). *An electronic environmental library project (Proposal to NSF)*. Available from: <http://http.cs.berkeley.edu/~wilensky/proj-html/proj-html.html>
- Woodruff, A. G., & Plaunt, C. (1994a). *Automated geographic indexing of text documents* (Sequoia 2000 Technical Report 94/41). Berkeley, CA: University of California, EECS.
- Woodruff, A. G., & Plaunt, C. (1994b). GIPSY: Geo-referenced Information Processing System. *Journal of the American Society for Information Science*, 45(9), 645-655.

LINDA L. HILL

Spatial Access to, and Display of, Global Change Data: Avenues for Libraries

Geospatial retrieval systems are considered in the framework of the U.S. Global Change Data and Information System (GCDIS), a cooperative effort of a set of federal agencies that have vast amounts of data and information relevant to the global change user community. The characteristics of an ideal geospatial retrieval system are proposed and five existing systems are described as examples of the types of spatial retrieval available today. Existing systems fall far short of the ideal system envisioned but show exciting promise. Librarians can participate in the evaluation of these systems and in the evolving development of GCDIS through the Library Information Subgroup of the Global Change Data Management Working Group.

INTRODUCTION

Spatial access to, and display of, global change data is addressed here in the context of the United States Global Change Research Program (USGCRP) and, in particular, of the Global Change Data and Information System (GCDIS). The USGCRP was established to observe, understand, and predict global change and to make its results available for use in policy matters. The goals of the data management activities of the USGCRP are to archive, preserve, and make data available in a useful form now and in the future for researchers and any other groups who can benefit from access to these data. The challenge of the data management program is to handle the massive amounts of highly diverse data and information generated and used by global change research activities. These data, and accompanying information products, reflect multidisciplinary global change issues that include human and natural forcing factors, ecological change, biodiversity, human interactions, and comprehensive assessments.

The federal agencies involved in the USGCRP have cooperated to organize the Global Change Data and Information System. The GCDIS builds on existing and planned agency-mission-responsive data and information systems that have information relevant to the global change user community and links these systems to each other and to users. The enhanced interoperability of the agency systems is achieved through standards, common approaches, technology sharing, and data policy

coordination. The design focuses on providing service for a diverse community of global change users—researchers, policymakers, educators, private industry, and the public. The objective is to aid the user community in learning what data and information are available, in having the key holdings available in useful forms with ready access, and in being assured of their quality and continued availability. Multiple levels of access and service are included in GCDIS, from phone and mail and fax to high bandwidth transfer of image data. The strategy for this initiative is described in the *U.S. Global Change Data and Information System Implementation Plan* (Committee on the Environment and Natural Resources, 1994). Other key documents from this effort are included in the references at the end of this discussion.

The Global Change Data Management Working Group (GCDMWG)¹ is building the GCDIS. Agencies participating in GCDIS are identifying their resources and developing coordinated ways for a wider audience than the primary research community to find and use their data and information. The GCDIS is being developed and implemented by the participating agencies with a minimum of additional infrastructure. The first interface effort was the beta version of the GCDIS gopher, registered in the Spring of 1994. In the Spring of 1995, the gopher is being revised on the basis of the “lessons learned” from the use of the beta version, and a web home page is being developed. These developments are being done through these agencies: the gopher is mounted at the National Oceanic and Atmospheric Administration (NOAA) by the Environmental Services Division Information Management staff, and NASA staff has contributed to its design; the DOE’s Carbon Dioxide Information Analysis Center (CDIAC) staff is developing the home page; the developments are being coordinated through the Executive Secretariat of the GCDMWG that consists of only two people: the Executive Secretary Les Meredith and myself. The working group itself is structured as follows:

- A Contacts Group guides the overall coordination effort for data management and three subgroups. It consists of representatives from the participating agencies. A Principals Group, which is the chief agency representative to the Working Group, is convened for budgetary decisions. There is no direct funding for the GCDIS effort. All activities are supported through contributions from the agencies. The Contacts Group is chaired by Dixon Butler of NASA’s Mission to Planet Earth Program.
- The Content Subgroup’s responsibility is to coordinate efforts to identify and set priorities for GCDIS data and information and related information products to resolve user needs, identify information gaps, encourage the participation of researchers in GCDIS, serve as a point

of contact for users that need special products, and coordinate the establishment of data and information quality assessment processes. The Content Subgroup is chaired by Michael Farrell of the Department of Energy.

- The Access Subgroup “strives to create a seamless search, browse, order, and delivery of global change data and information from many agencies” (from draft GCDIS implementation plan, March 1995), promotes the use of the Z39.50 protocol and emerging Internet tools, and is taking an active role in bringing together the players in other related federal initiatives: the Federal Geographic Data Committee (FGDC), the National Environmental Data Inventory (NEDI), and the Government Information Locator Service (GILS). The development of the GCDIS gopher and the home page is being done through the Access Subgroup. The Global Change Research Information Office (GCRIIO), which provides help desk support to GCDIS, is linked to this subgroup. The Access Subgroup is chaired by Tom Mace of the Environmental Protection Agency.
- The third subgroup is the Library Information Subgroup (LIS). GCDIS may be the only high level federal government interagency data activity that specifically brings library and information center resources and expertise to the table at this level. The responsibility of the Library Information Subgroup is to support the Working Group efforts by:
 - A. building an infrastructure of libraries and librarians for GCDIS implementation;
 - B. evaluating GCDIS from a library user’s perspective and providing user needs analysis;
 - C. linking data resources to information resources for knowledge management;
 - D. promoting GCDIS to libraries and developing approaches to user education;
 - E. advising on data and information processing standards and systems from the library perspective (from draft LIS implementation plan, March 1995).²

The Library Information Subgroup is chaired by Roberta Rand of the National Agricultural Library.

Last year each subgroup began sponsored pilot projects: the Contents Subgroup is investigating methodologies for determining priorities of research interest and thus content acquisition using the trace gas research community as the testbed. The Access Subgroup is sponsoring the Assisted Search for Knowledge (ASK) project that is developing a client-server search, retrieval, and use system based on a semantic network

search engine. This will be discussed later here because the ASK project has a geographic information system component. The Library Information Subgroup is sponsoring a GCDIS evaluation project in the state of Virginia that involves public school (elementary, upper elementary, and high school), community college, university, public library, and community groups in the use of GCDIS as it exists and involves them in the further design and enhancement of the system. This is known as the Library Access, Search, and Retrieval (LASR) Pilot Project and is being led by the Computer Sciences Department of the University of Virginia.

The data and information resources that are within the purview of the GCDIS effort include both those resulting from focused Global Change Research Project (GCRP)-funded research and the data and information resulting from other agency efforts. The potential resources to be tapped include (from the draft implementation plans of the participating agencies submitted in the Fall of 1994):

- terabytes of data derived from Environmental Protection Agency (EPA) regulatory activities;
- more than two dozen data and information service nodes in the Department of Commerce, including the satellite and in situ data of the National Oceanic and Atmospheric Administration; the holdings of the NOAA Central Library; the statistical, economic, and demographic information of the Bureau of the Census; and the technical, engineering, and business information available through the National Technical Information Service (NTIS);
- data and information from the Department of Energy's Atmospheric Radiation Measurement (ARM) Archive, the Carbon Dioxide Information Analysis Center, the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC), their Office of Scientific and Technical Information (OSTI), the Energy Information Administration (EIA), and the Energy Efficiency and Renewable Energy Network (EREN);
- data archives from the Department of Agriculture's Agricultural Research Service (ARS); Cooperative State Research, Education and Extension Service (CSREES); their Economic Research Service (ERS); the Forest Service (FS); and the Terrestrial Ecosystems Regional Research and Analysis (TERRA) Laboratory as well as the holdings of the National Agricultural Library (NAL) and the World Agricultural Outlook Board (WAOB);
- the terabyte data archives of the National Science Foundation's National Center for Atmospheric Research (NCAR), including the NCAR library holdings, and the Long-Term Ecological Research Network;
- the data servers of three Department of Interior programs: The Global Change Research Program Data Server, the United States GS Earth

Resources Observation Systems (EROS) Data Center, and the National Biological Service (NBS) Global Change Research Program Data Management Center as well as additional data centers and data systems that archive many types of geospatial data (e.g., water quality measurements, forest inventory plots, soil sampling, wetland maps, digital cartographic data files, census counts, and timber inventories);

- the Global Change Master Directory (GCMD) and the Earth Observation System Data and Information System (EOSDIS) from the National Aeronautics and Space Administration (NASA) as well as data and information from the Pathfinder program, the Crustal Dynamics Data Information System (CDDIS), and NASA's Scientific and Technical Information Program.

Most of the agencies, in addition to the resources listed here, have their own data-sharing agreements or links to state, regional, and worldwide resources. There are also other nodes such as NASA's Consortium for International Earth Science Information Network (CIESIN) which provides access to socioeconomic data and information as well as scientific data and information and the Global Change Research Information Office which provides user services to those accessing this wealth of information.

Each participating agency has agreed to a set of minimum requirements and has prepared its own GCDIS implementation plan. Each subgroup has also prepared an implementation plan targeting activities for the next three years.

This brief introduction is intended to give the reader a picture of the vast amount of national and international data and information resources within the purview of the GCDIS and the interagency structure established to provide coordinated access to these resources. This whole effort is not happening in isolation, by any means. The National Academy of Sciences reviews and approves the Working Group plans and through overlapping committee memberships, reciprocal presentations and discussions, and focused coordination meetings, the GCDIS effort is harmonized with similar programs. The GCDIS effort seeks to be an entry point to distributed global change data and information for diverse audiences and particularly to make the data and information accessible by and through libraries.

There are many substantive issues in relation to GCDIS: priorities for content acquisition; data management priorities; interoperability, accessibility; management of a distributed system; political and budgetary issues for a multiagency effort; resource discovery structures and mechanisms for a distributed system; metadata standards; meeting the needs of a broad range of users; and the international, multifformat, and interdisciplinary aspects of the system. The question that will be explored

here is focused on the extent to which GCDIS, in its present status, can provide access to the available data and information by geographic location through spatial access, display, and manipulation. Although the emphasis in GCDIS is on global data sets and information and many of the problems in this area require global data gathering (e.g., global warming, ozone depletion, El Niño, etc.), GCDIS data and information can often be linked to specific geographic areas through latitude and longitude coordinates and geographic place names.

Before illustrating the current status of spatial access to federal government global change data and information, some of the characteristics of an ideal spatial retrieval system for finding georeferenced information will be outlined.

SPATIAL ACCESS—WHAT WOULD WE REALLY LIKE TO BE ABLE TO DO?

As background, my primary interest is in information retrieval. The research that was done for my dissertation (Hill, 1990) compared the effectiveness of the retrieval of georeferenced information using word-based descriptions (place names) and spatial representations. The research was based on ninety-nine geoscience articles about the Mediterranean region, each having a sketch map of the research area. Spatial representations of the maps contained in the articles were used as the “true” representations of the geographic study areas. This was done with the GRASS GIS software³ and resulted in raster representations of the sketch maps in the articles. For every document in the set, the other ninety-eight documents were ranked by geographic similarity on the basis of their spatial overlap or distance away. Rankings between documents were also determined based on geographic terminology, both assigned descriptors and the text of the bibliographic citations and abstracts. Comparison between spatially-based rankings and word-based rankings were used to determine how successfully the words in the bibliographic records reflect the geographic relationships among the documents. The results showed that there was only a weak correlation between the rankings based on geographic words and the rankings based on spatial map-based representations.

The study revealed other confirmations of the ineffectiveness of words for the retrieval of georeferenced information. Using spatial overlap as a sign of relevance between one document and another, word-based retrieval was only able to average 50 percent recall and 41 percent precision—that is, only 50 percent of the relevant documents were identified (on average) and only 41 percent of the documents retrieved by geographic words were overlapping areas on average. These recall and precision figures are typical of other studies of the effectiveness of text-based retrieval; this

impreciseness of our retrieval tools is not limited to geographic terminology. But in the case of the geographic “aboutness” of documents, we do have an alternative—a spatial representation of the areas.

Some of the bibliographic records used in the study were from GeoRef^d and contained latitude and longitude coordinates to describe the study areas. There were obvious errors in a few of these due to such errors as reversing top and bottom latitudes or conversion errors when the records were loaded on the Orbit system. When seen graphically, the errors are obvious, but these errors are very difficult to detect when looking at a string of latitude and longitude values. The other problem with the system of assigning two latitudes and two longitudes to describe an area—still the dominant approach—is the impreciseness of this regular polygon parallel to the equator when describing an irregular area. However, the rankings based on latitude and longitude representations did have a higher correlation to the spatial representations than the words in the study.

My ideal system for finding and retrieving georeferenced information will have text documents, data sets, maps, photographs, and all other forms of information that are geocoded to represent the geographic coverage areas of the items, formatted in accordance with widely accepted interoperable metadata standards. The users of this system would be able to retrieve all forms of information that relate to particular geographic areas from distributed databases.

The geocoding describes geographic areas—not just a latitude and longitude point but the representation of an area of coverage or study. The areal representation should closely match the study area of the document or data set and not be grossly generalized. For example, a data set or technical report covering the Gulf Coast of Florida should not be represented by a box that also takes in most of the Gulf of Mexico. A river valley should not be represented by a point at the mouth of the river nor by a box that greatly exceeds the environs of the river valley. The representation of the area should be close enough to the actual geographic area for appropriate differentiation for information retrieval. This does not require the exactness of a map representation but, in an ideal system, geographic representation would not be limited to either single points representing the geographic center of an area or to regular polygons that are parallel to the equator.

The spatial query capability in the ideal system would be imbedded in a retrieval system that supports other types of query parameters: free-text words and phrases from the documents and metadata; subject headings and descriptors; dates and date ranges; authors; organizational affiliations and sponsorship; format; type of information; and other technical details. The spatial query capability would present the user with a map of

the world. Through pointing and zooming, the user zeros in on the area of interest and draws a shape that surrounds the area of interest. The user is not limited to a rectangular shape but can instead outline the Gulf Coast of Florida or the Mississippi River valley; an oil producing region or a fault zone; a forest tract or a park; a biological zone or an ocean current. If simple geographic names or census tract numbers or political entities or other easily named areas will suffice for the query, then the user can simply enter these names with the assurance that the system can, if requested to, translate the place names into spatial definitions and expand the search to those items in the database that may not have those specific names attached to them but that include the area nevertheless.

To bridge the gap between data that are geocoded with latitude and longitude coordinates (or a scheme that can be translated to latitude/longitude descriptions) and documents that have their geographic "aboutness" described with place names, there needs to be a translation scheme. This would be a database that defines place names spatially with latitude and longitude coordinates so that the information-retrieval system can prompt users for place names that are in a designated geographic search area or expand a search that includes place names to their spatial equivalents. In this way, users of the information-retrieval system can use either coordinates or place names and have the option of expanding their search to the other form of description. There is some preliminary work being done in this area and a great deal more needs to be done.

The matching mechanisms for this ideal retrieval system will compare the user's query area with the spatial coverage areas of the items in the database (or databases) and return a ranked list based on a metric, such as a similarity measure.

- For overlapping areas (i.e., the query area and the item area overlap), the area of the overlap times 2 would be divided by the sum of the query area plus the area of the item that was matched. This process adjusts for widely differing sizes between the query and the target item in the database as well as the amount of overlap. The highest similarity value comes from the largest overlap and the most similar sizes. A global data set would have a low similarity to the Florida Gulf Coast even though it does contain it with this scheme. All database items with areas overlapping the query area would be ranked by the system and returned in descending order by the strength of the similarity.
- Database items whose geographic areas share a common boundary with the query area but do not overlap it would have a similarity value of zero.
- Database items that do not overlap the query area and do not have a common boundary would be related to the query area by a negative number representing the distance between the closest boundaries.

The user would be allowed to set a similarity cut off value so that only overlapping areas are retrieved (i.e., similarity greater than zero) or only database items with a "significant" amount of overlap.

Probably the ideal system will provide various ways of limiting or filtering retrieved sets of data and information by spatial extent. A metric could be used that indicates the percentage of the query area covered by the database items so that complete coverage could be distinguished from partial coverage. Doug Nebert of the USGS and Project Alexandria at the University of California, Santa Barbara, is also considering other ways of giving the user a sense of the spatial extent of the database items in relation to the query area. One idea is to compute the “square degrees” of an area by multiplying the latitude and longitude dimensions of spatial descriptions. With this artificial measure, areas could be divided into rough sets such as “local,” “regional,” “continental,” and “global” and the user given a choice of the extent of coverage she or he wants to have in the retrieved set.

The system would return a ranked list of the database items that meet all of the search criteria with the geographic component of the search expressed in spatial terms. Exactly how this is to be done has not received much attention, as far as I know. The user would now like to ask the system to display the items by their geographic areas on the map used for the query and show the types of items by a visual key: data sets, images, full text, metadata only, books, articles, technical reports, maps, photographs, etc. From this display, items can be chosen for review and selection—for downloading or ordering.

Now with the data in hand, the user will need further system help to manipulate the data sets and images selected. This will involve browsing snapshots of the images (or progressive displays of the images that the user can stop when she or he has seen enough⁵) before downloading them. Full documentation for the data should be available so that it is clear how the data are registered and encoded; the circumstances under which the data were obtained; the instruments used; the quality control applied; etc. Eventually, the user may want to display and manipulate the data sets with a Geographic Information System (GIS) and correlate data from different sources.

In summary, the ideal system will present a graphic spatial query mode for geographically related searches; it will return a ranked list of relevant items based on a similarity measure or other metric; it will display the resulting “hits” on a map with proper identification so that items can be reviewed for selection; and it will provide tools for browsing images and manipulating data sets.

FIVE EXAMPLE SYSTEMS FROM FEDERAL AGENCIES WITH SPATIAL RETRIEVAL

The Global Change Data and Information System provides an initial starting point to a rich, diverse, and distributed set of scientific and social data and information resources. The major contributors to GCDIS were introduced earlier in this discussion. Participation is not limited to these

seven agencies, however. Any federal agency with data and information that can contribute to an understanding of global change is being encouraged to participate in GCDIS, and the agencies' international, state, local, regional resources will also be incorporated. For the purpose of representing the spatial access, display, and manipulation capabilities of this diverse group, the following systems will be highlighted:

- The GC-ASK Project
- The Global Change Master Directory
- NASA's EOSDIS Version 0 Information Management System
- NOAA's National Geophysical Data Center/World Data Center—A Marine Geology Inventory Online Search System
- U.S. Geological Survey Node of the National Geospatial Data Clearinghouse, a component of the National Spatial Data Infrastructure (NSDI)

These systems have been chosen as representative of the types of geospatial information retrieval available today. At the end of this discussion there is a list of World Wide Web sites (Appendix A) of interest for geospatial data and information. Again, this is not an exhaustive list but offers several starting points to an exploration of the Web.

The GC-ASK Project

The Global Change Assisted Search for Knowledge (GC-ASK) is a one-year technology development project of the GCDMWG's Access Subgroup. The design, built on commercial off-the-shelf (COTS) products, incorporates the semantic network retrieval engine of ConQuest⁶ into a client/server architecture to access distributed databases of all kinds. Plans include the integration of GIS software from E-systems⁷ called OASIS (Open Architecture for Scientific Information Systems). There are four prototypes planned in the one-year project. The first prototype demonstrated the ability of the system to search distributed databases that had been indexed by ConQuest software. The second prototype includes a Graphical User Interface, Z39.50 compliance, and additional terminology collections (i.e., National Institutes of Health's Unified Medical Language System [UMLS], Defense Technical Information Center [DTIC] thesaurus, and the NASA Thesaurus). Its debut will be on Earth Day, April 22nd, on the Mall in Washington, D.C.

The third prototype is scheduled to include GIS capability, additional interfaces for multiple user classes, a consistent data presentation model, and connections to non-native (non-ConQuest) search engines to demonstrate the capability to link to and effectively use the functionality of existing search engines of participating agencies. The fourth and final prototype is scheduled to provide the ability to perform assisted searches simultaneously across multiple databases—including non-ConQuest

search engines—and includes retrospective searching, real-time profiling, and on-disk (CD-ROM) product searching. The project will be completed in October 1995.

The GIS component is being contributed to the GC-ASK project by E-systems. They will use their Open Architecture for Scientific Information Systems software product which builds on common industry standards and commercially available software. Core OASIS functions include georeferenced data management using a wide variety of data sources and formats and support for data dissemination, handling, and storage and exploitation for research and analysis. These functions stem from the integration of a geographic information system with image processing and a relational database management system (RDBMS) with a mechanism for incrementally adding data visualization tools. Depending on the capabilities of the host workstation, the OASIS system can display data in a variety of formats and combinations.

The GIS software used in OASIS is Genemap, an open standards-based vector GIS. The functionality envisioned for GC-ASK is a map-based interface that permits zooming and selection of a polygon or point representing the area of interest. The results are displayed graphically, superimposed over the query area as well as in a tabular listing of metadata for the retrieved data sets or documents. Exact implementation of GIS capability in GC-ASK has not been determined at the time that this presentation was written.

GC-ASK is a technology development project—a proof of concept. If it is successful and it is adopted by the GCDIS agencies, users will be able to access both distributed data and information sources with spatial and natural language queries. Spatial access depends on the database objects having geocoding, of course, which is not currently available for the textual records and documents. Another handicap for spatial access to text in the GC-ASK prototype is that the current terminology bases used by ConQuest do not include a geographical thesaurus of place names and their relationships.

The Global Change Master Directory (NASA)

NASA's Global Change Master Directory <<http://gcmd.gsfc.nasa.gov>> is an international multidisciplinary directory of metadata records for scientific data sets, including satellite and ground observational data, chronological climatological data, and data from global and regional field experiments. It is a major component of GCDIS.

Data sets on individual scientific workstations as well as those in agency data centers are included in the GCMD. The GCMD data descriptions include NASA, NOAA, National Center for Atmospheric Research, USGS, DOE's Carbon Dioxide Information Analysis Center, and EPA data sets, along with entries from universities and research centers. The GCMD

also contains descriptions of data held outside the United States through the International Directory Network (IDN). The GCMD is duplicated in three non-U.S. sites (Italy, Japan, and Canada) and updated every two weeks.

The GCMD database contains approximately 3,000 high-level data set descriptions that give the user basic information on the data sets and points of contact. Through a LINK command, an automatic connection to the site of the data holding can be created, linking directly to external systems for more information, browsing, and even data ordering. These sites currently include the NASA Space Shuttle Earth Observation Program (SSEOP), NOAA National Climatic Data Center (NCDC), USGS Earth Resources Observation Systems Data Center, Global Land Information System (GLIS), CNES SPOT Image Catalog, Earth Observation Satellite Data Inventory Service (SINFONIA), and the University of Rhode Island AVHRR Archive. Future links will include the Earth Observing System Data and Information System and the Consortium for International Earth Science Information Network. The remaining part of the GCMD database encompasses supplemental information on data centers/systems, campaign/projects, spacecraft, and sensors.

GCMD data sets are described in the Data Interchange Format (DIF), which is being modified to comply with the Federal Geographic Data Committee metadata standard. The FGDC metadata standard and the DIF include the following locational data elements:

West_Bounding_Coordinate
East_Bounding_Coordinate
North_Bounding_Coordinate
South_Bounding_Coordinate

These four bounding coordinates can be used to retrieve data sets from the GCMD when the "Query for DIFs" option is chosen from the WWW home page. The other search criteria, such as "Location Keyword," are supported by the lists of controlled vocabulary permitted in the field—the "Valid's."

In the WWW interface, there are boxes provided for entering the "Southwestern Coordinate (lat, lon)" and the "Northeastern Coordinate (lat, lon)." The GCMD WWW interface includes a choice of including or excluding the global data sets from the retrieval. The matching algorithms will select any area that overlaps, encloses, or is enclosed by the target area. GCMD can also be accessed through a character interface and an X-Windows interface.

The system response to a query is in the form of a listing of metadata records. The user must take additional steps to obtain access to the data, and contact points are included for this purpose. The GCMD is a direc-

tory level database, meaning that one metadata record can represent a set of data. The data set as a whole may have global coverage, and that is the only geographic level that is provided at the GCMD level. The more detailed level of data coverage—the inventory level—is available from the data center archives.

The WWW interface to the GCMD also includes the option to search the database with the WAIS software. Future plans include improving geographical searches, implementing spatial WAIS, and migrating the system to a “distributed” architecture.

EOSDIS Version 0 Information Management System (NASA)

The Earth Observing System (EOS), part of NASA’s Mission to Planet Earth, is NASA’s major contribution to the Global Change Research Program. The Data Information System component, EOSDIS, is being designed as a distributed system to support archiving and distribution of data at multiple data centers. The Distributed Active Archive Centers (DAACs) are connected by an Information Management System (IMS) which provides an interface for “one stop shopping” for earth science data, allowing users to search for and order data from multiple data centers in a single session.

Version 0 of EOSDIS, released in September 1994, is a prototype system <http://harp.gsfc.nasa.gov:1729/eosdis_documents/eosdis_home.html>. The interface is available through both graphical and character user interfaces to accommodate a variety of user environments ranging from simple VT100 terminals to sophisticated graphical workstations. It allows the user access to information and metadata through three levels or types of information—directory, guide, and inventory searches. The directory provides concise high-level information about data sets from any point in the system. The guide provides detailed descriptions about data sets, platforms, sensors, projects, and data centers. The inventory search function provides descriptions of specific observations or collections of observations of data (granules) that are available for request from a data archive. Version 0 can be used through any of the NASA DAACs and all of the connections to the DAACs are present on the Version 0 home page. Other functions provided include:

- A Product Request allows users to view information pertaining to orderable data products and then construct a request which is forwarded to the relevant archive for order processing.
- The Coverage Map, available in the graphical version, is a two-dimensional graphical representation of the geographic coverage of selected inventory granules. It displays the Earth in an orthographic projection.

- The Browse function allows a user to locate and retrieve reduced resolution images as an aid to data selection. The user may either view the image in the graphical version IMS interface or have it staged for FTP pick-up.

The Version 0 home page provides links to other home pages that describe various EOSDIS DAACs and cooperative data centers. User support contact information is provided. Also available is a “glossary of terms,” an “acronym list,” and a “list of known problems.”

The geographic search is supported in multiple ways through the X-Windows interface: by selecting from a map, describing a rectangle, by describing four corners of a polygon that do not have right angles and are not parallel to the equator, by describing a point and the height and width of the area around the point, by describing just a point reference, or by global searches. Search results can be limited to only those that are global but it is not possible to eliminate global data sets from the results. This means that a search for any small area will retrieve all global data sets.

Data sets can be selected from the resulting lists of “hits” from each data center based on their descriptions, based on their coverage areas which can be displayed on map projections, or on the basis of browse images when they are available. Orders for the data sets can be entered online.

The EOSDIS IMS software continues to be developed and will be available in an updated version soon. This client software will be available for distribution for those that want to use it locally.

The National Geophysical Data Center (NOAA)

The National Geophysical Data Center (NGDC) <<http://www.ngdc.noaa.gov/>> manages environmental data in the fields of solar-terrestrial physics, solid earth geophysics, marine geology and geophysics, paleoclimatology, and glaciology (snow and ice). In each of these fields it also operates a World Data Center (WDC) discipline center (from the NGDC home page). The home page points to several sites of interest to a GIS audience. For an illustration of “GIS Techniques Using NOAA Data,” choose “Solid Earth Geophysics” and then “GIS Applications using NOAA Data” <<http://www.ngdc.noaa.gov/seg/globsys/gisdes.html>>. For an example of spatial searching, choose “Marine Geology and Geophysics” <<http://www.ngdc.noaa.gov/mgg/mggd.html>> and then “Online Searches” and then “SEARCH GEOLIN” <<http://www.ngdc.noaa.gov/mgg/geolin/gsearch.HTML>>.

“The GEOLIN system contains an inventory of all marine geological data sets and reports available from NGDC in digital or analog form and

supports spatial searching via the 'latest' forms interface of X-Mosaic or Mosaic for Windows" (from the home page). The GEOLIN search uses Mosaic forms to launch an experimental version of cross-platform libraries developed at NGDC in cooperation with DataWare Technologies, Inc. of Boston, Massachusetts, and Boulder, Colorado, based on the Record ReferenceBook libraries owned by DataWare Technologies, Inc. (from the GEOLIN introductory page).

In this forms-based spatial searching interface, users are prompted for upper and lower latitude and left and right longitude. The minus sign is used for southern latitudes and western longitudes, a common device. All overlapping areas are retrieved no matter how broad the coverage.

Also from the National Geophysical Data Center in Boulder, Colorado <http://www.ngdc.noaa.gov/seg/globsys/global_c.html> are global environmental data sets on CD-ROM produced by the Global Ecosystems Database Project, conducted by NGDC and the U.S. Environmental Protection Agency's Environmental Research Laboratory in Corvallis, Oregon. The databases contain raster gridded map layers registered to a common latitude-longitude base. Each file has been inspected for optimal quality and usability for analysis based on the current state of data development. Parameters have been chosen for their potential use in integrated studies of the global environment. Individual data layers have been contributed by many scientific laboratories and individuals. More information about these CD-ROMs is included in Appendix C of this paper.

The NOAA Data Set Catalog <<http://www.esdim.noaa.gov/NOAA-Catalog/NOAA-Catalog.html>> provides word searching, phrase and root searching, Boolean and field-based searching of its metadata records, and gives a generous set of information to help users format and conduct their searches. They do not have a spatial component currently but include "latitude/longitude searches" in their list of future plans.

National Geospatial Data Clearinghouse (USGS)

The USGS node of the National Geospatial Data Clearinghouse is a component of the National Spatial Data Infrastructure <<http://nsdi.usgs.gov/nsdi/index.html>>. Sets of USGS geospatial data can be searched by keyword or spatial extent using SpatialWAIS. Data sets are described using the Federal Geographic Data Committee Content Standards for Digital Geospatial Metadata.

Browsing and searching capabilities are available. By using the search option, you can identify your interest by keyword (a WAIS search) or you can identify an area by its latitude and longitude boundaries using the forms interface to WAIS. Another option is to choose a state or states as the area.

Once you find a data set you want, its metadata files contain instructions for obtaining the data set. Some smaller data sets are available online at no cost. For others, you may need to set up an account to pay a small distribution fee. A few data sets are not online, and you will be instructed how to order a compact disk or a tape.

Many USGS data products are available in the Spatial Data Transfer Standard (SDTS) format, a new federal standard to ensure data compatibility, or in the older Digital Line Graph (DLG) format. Others are in more specialized formats, which are described in their metadata. Whenever you retrieve a data set, it is important to retrieve the metadata as well, since the metadata provides important information for using the data set.

The ranking algorithm that is operating in this system is based only on the words entered for the WAIS search—the geographic parameters are not incorporated. Therefore, if you enter a search composed only of spatial parameters, the ranking that is returned is purely arbitrary.

The USGS is moving to the I-Site software from the Clearinghouse for Networked Information Discovery and Retrieval (CNIDR) (Nebert, 1995). I-Site is based on version 2 of the Z39.50 protocol with the incorporation of some elements from version 3; it has field-based search capability as well as the means by which to call other search software. A description of I-Site can be located at <http://vinca.cnidr.org/software/Isite/Isite.html>.

SUMMING UP—WHAT IS THE STATUS OF SPATIAL DISCOVERY AND DISPLAY TOOLS FOR GLOBAL CHANGE DATA?

The spatial access tools for the Internet environment are in the developmental stages. They show exciting promise for the near future as additional tools for network discovery of data relevant to the study of a particular geographic area. More visual map-based interfaces are appearing that support the representation of areas. Most systems, however, are still limited to the representation of regular polygons or points. Matching algorithms to return a ranked list on the basis of geographic similarity are not being used—at least not in the systems that I reviewed.

Spatial access to textual information along with data sets is an area that needs more emphasis through research and development to build translations between the geographic text and geocoded representations. There are several efforts in this direction. The GeoWeb Project is being developed by Brandon Plewe of the State University of New York at Buffalo (plewe@acsu.buffalo.edu) and Chris Stuber of the U.S. Census Bureau (Chris.Stuber@census.gov) (Plewe, 1994). The concept of GeoWeb

includes a gazetteer interface to expand geographic terminology to a latitude/longitude point reference that can be used for searching. The second is the GIPSY project by Allison Woodruff and Christian Plaunt (1994) at the University of California, Berkeley, which processes the natural language of the document with a "spatial reasoning" algorithm "to approximate statistically the geoposition being referenced in the text." GIPSY uses not only place name terms from the text but also terms that describe geologic features, climate, land use, habitats, and area size and reference terminology sets such as the USGS Geographic Names Information System and the USGS Geographic Information Retrieval and Analysis System (GIRAS). In addition, the NSF/ARPA/NASA digital library project at the University of California, Berkeley, is planning to develop a gazetteer-spatial link for their project.

There is potential for users seeking georeferenced information to discover and obtain the information they need. Fortunately, many free sources of information exist—many open access systems are being built. However, most users are going to need help—the tools are not standardized enough, the systems are not interoperable enough, the coverage is not complete enough, and there is too much to keep up with in this rapidly changing environment. This help should be available through library services that specialize in geographic information as well as specific governmental offices.

GCDIS, in particular, provides an avenue for librarians to participate in the building and expansion of the systems to access global change data and information, including spatial data and information. Through the Library Information Subgroup, librarians can directly participate in evaluation projects and voice their needs and ideas in an effective way. In particular, librarians are invited to participate in the evaluation of the GC-ASK retrieval software engineering project, in the evaluation of the GCDIS gopher and home page, and in the evaluation of the systems available through the participating federal agencies.

APPENDIX A. INTERNET SITES FOR THE SYSTEMS DISCUSSED AND RELATED SITES

Board of Geographic Names

http://toponym.dma.gov/check_login.html

CIESIN

<http://www.ciesin.org>

EOSDIS

http://harp.gsfc.nasa.gov:1729/eosdis_documents/eosdis_home.html

Fish and Wildlife Service Hotlist

<http://www.fws.gov/hotlist.html#Geospatial>

GC-ASK (Global Change - Assisted Search for Knowledge Project)

telnet://esdim2.edim.noaa.gov

Login id is "conquest"

Password is "conquest11"

For more information about the project, send e-mail to
gcdis.lis@earth.usgcrp.gov.

GCDIS (Global Change Data and Information System)

[gopher://gopher.gcdis.usgcrp.gov](http://gopher.gcdis.usgcrp.gov)

If login required: "gopher"

<http://www.gcdis.usgcrp.gov> (under construction)

E-mail: gcdis.lis@earth.usgcrp.gov

GCMD (Global Change Master Directory)

<http://gcmd.gsfc.nasa.gov>

GCRI (Global Change Research Information Office)

<http://www.gcric.org>

[gopher://gopher.gcric.org](http://gopher.gcric.org)

E-mail: help@gcric.org

GEONET-L (Geoscience Librarians & Information Specialists)

geonet-l@iubvm.ucs.indiana.edu

GeoWeb Project (description included in Appendix C)

<http://wings.buffalo.edu/geoweb/general.html>

Ionia "1 km AVHRR Global Land Data Set" Net-Browser (European Space Agency, Italy)

<http://shark1.esrin.esa.it/>

LASR Project (Library Access, Search, and Retrieval Project to evaluate GCDIS)

<http://juliet.cs.virginia.edu/lasr/>

Marc's Favorite Internet Sites (Marc E. Berryman, marc.berryman@dir.state.tx.us)

<http://www.texas.gov/DIR/people/mab.html>

National Geospatial Data Clearinghouse. U.S. Geological Survey Node

<http://nsdi.usgs.gov/nsdi/index.html>

NOAA Marine Geology & Geophysics On-line Searches

<http://www.ngcd.noaa.gov/mgg/geolsys.html>

HILL/SPATIAL ACCESS AND DISPLAY OF GLOBAL DATA

Sol Katz List of Spatial Geographic Search Sites

<ftp://ftp.blm.gov/pub/gis/nsdi.html>

TMS Experimental Browser (description included in Appendix C)

<http://tiger.census.gov/>

APPENDIX B. ACRONYMS

ARM	Atmospheric Radiation Measurement
ARS	Agricultural Research Service, USDA
AVHRR	Advanced Very High Resolution Radiometer
CDIAC	Carbon Dioxide Information Analysis Center, DOE
CIESIN	Consortium for International Earth Science Information Network, NASA
CSREES	Cooperative State Research, Education and Extension Service, USDA
CDDIS	Crustal Dynamics Data Information System, NASA
DAAC	Distributed Active Archive Center
DIF	Directory Interchange Format, GCMD
DOE	Department of Energy
DOI	Department of Interior
DTIC	Defense Technical Information Center, DOD
EIA	Energy Information Administration, DOE
EOSDIS	Earth Observation System Data and Information System, NASA
EPA	Environmental Protection Agency
EREN	Energy Efficiency and Renewable Energy Network, DOE
EROS	Earth Resources Observation Systems, USGS
ERS	Economic Research Service, USDA
ESA	European Space Agency
ESDIS	Earth Science Data and Information System
FGDC	Federal Geographic Data Committee
FS	Forest Service, USDA
GC-ASK	Global Change Assisted Search for Knowledge Project
GCDIS	Global Change Data and Information System
GCDMWG	Global Change Data Management Working Group
GCMD	Global Change Master Directory
GCRIO	Global Change Research Information Office
GCRP	Global Change Research Program
GILS	Government Information Locator Service
GIS	Geographic Information System
GUI	Graphical User Interface
IWGDMGC	Interagency Working Group on Data Management for Global Change (previous name for GCDMWG)
LASR	Library, Access, Search, and Retrieval Project, LIS
LIS	Library Information Subgroup, GCDMWG
NAL	National Agricultural Library, USDA
NASA	National Aeronautics and Space Administration
NBS	National Biological Service, DOI
NCAR	National Center for Atmospheric Research, NSF
NEDI	National Environmental Data Inventory
NGDC	National Geophysical Data Center
NOAA	National Oceanic and Atmospheric Administration, Department of Commerce
NODIS	NASA Online Data and Information System
NSF	National Science Foundation
NSDI	National Spatial Data Infrastructure

HILL/SPATIAL ACCESS AND DISPLAY OF GLOBAL DATA

NSSDC	NASA Space Science Data Center
NTIS	National Technical Information Service (NTIS)
OASIS	Open Architecture for Scientific Information Systems, E-Systems
ORNL	Oak Ridge National Laboratory, DOE
OSTI	Office of Scientific and Technical Information, DOE
TERRA	Terrestrial Ecosystems Regional Research and Analysis Laboratory, USDA
UMLS	Unified Medical Language System
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
WAOB	World Agricultural Outlook Board, USDA
WWW	World Wide Web

APPENDIX C. NATIONAL GEOPHYSICAL DATA CENTER, BOULDER CO. GLOBAL CHANGE CD-ROMS (ADAPTED FROM HOME PAGE)

Also from the National Geophysical Data Center in Boulder, Colorado <http://www.ngdc.noaa.gov/seg/globsys/global_c.html>, are global environmental data sets on CD-ROM produced by the Global Ecosystems Database Project, conducted by NGDC and the U.S. Environmental Protection Agency's Environmental Research Laboratory, in Corvallis, Oregon. The databases contain raster gridded map layers registered to a common latitude-longitude base. Each file has been inspected for optimal quality and usability for analysis given the current state of data development. Parameters have been chosen for their potential use in integrated studies of the global environment. Individual data layers have been contributed by many scientific laboratories and individuals.

"Disc A" includes selected data on the global environment, such as ecosystems, land use, wetlands, vegetation (including satellite-derived vegetation index), climate, topography, and soils. These data are on a range of compatible grids, from 2 degrees to 2 minutes. Vector data for coastlines and other features are also provided.

"Disc A," in beta test since last year, is readable on IBM DOS-compatible machines, Apple Macintoshes, UNIX workstations, and other computers that support the ISO 9660 standard. Features of the Global Ecosystems Data on CD-ROM:

- *Integration of Data for Multivariate Analysis.* The data have undergone extensive quality control and preparation of documentation (300+ pages) for improved data synergy and usefulness for integrated multivariate spatial analysis. Several new techniques have been developed in this process. This process included format and grid conversion, registration and/or reprojection, creating documentation as needed, etc., while preserving the original nature and content of the data.
- *Scientific Peer Review of Data.* From NGDC's network of colleagues, cooperative projects with the International Geosphere-Biosphere Programme and the U.S. Environmental Protection Agency, over 100 scientists have reviewed the data. This review, based on prototypes since 1989 and a full beta test of a prototype CD-ROM in 1991-1992, has resulted in significant improvements to design, documentation, and content. Peer review comments are summarized with the database. This process continues, supporting future updates.
- *Public Domain, Accessible.* The data are completely within the public domain, per agreement by authors of each data set. All "value-added" work performed at NGDC in integrating the data is also in the public domain.
- *Data and Software Support Research, Education, Awareness—Optimized for use on Geographic Information Systems.* We provide software to browse, visualize, and select appropriate portions of the database for your work. Primarily, however, the database is designed to interface with many statistical, image analysis, and geographic information systems. Data and sample setups are provided in two highly accessible geographic information systems—GRASS and IDRISI—to show how they can be used in other GISs. The structure of the database has demonstrated its adaptability by passing several tests by software vendors for quick and easy importation into their systems.

HILL/SPATIAL ACCESS AND DISPLAY OF GLOBAL DATA

- *Provide Feedback to Authors of Data Sets.* The process of peer review and integration includes continued interaction with authors of individual data sets. Data enhancements and corrections are made cooperatively with creators of the data sets. The feedback that we obtain in our normal duties as a Data Center also helps to provide ideas and resources for improving data sets.

There is also a beta version of "Disc B" which contains:

Global Vegetation Index data produced at Chiba University, Japan
Weekly sea surface temperature
Derived snow cover and snow depth data
Ecoregions data
Improved urban coverage data
Elevation and bathymetry
Forest practices database
EPA climate database and doubled CO₂ predictions for the US
Potential natural vegetation of the conterminous US
US boundaries from Micro World Data Bank II
EPA rice climatology—UV-B irradiance and agroclimatology for South, Southeast, and East Asia

Other global and regional data sets available from NGDC:

- *Experimental Calibrated Global Vegetation Index on CD-ROM.* Data from NOAA's Advanced Very High Resolution Radiometer. Originally designed and produced by Kevin Gallo of the Office of Research and Applications of NOAA's National Environmental Satellite, Data and Information Service. Please request flier SE-2008 for more information about this data set.
- *Global Change Education Diskette Project.* Originally developed for the International Geosphere-Biosphere Programme, this is an integrated Global Change Data Base for Africa, with accompanying documentation and a workbook of exercises designed for classroom and self-instruction in using Geographic Information Systems to explore the global or continental environment. Please request flier SE-2007 for more information about this data set.

How to Order:

The price for the Global Ecosystems Data on CD-ROM, with browse and visualization software and extensive documentation is \$101 (product number 1016-A27-001). Data contributors and academic researchers should contact NGDC for information about obtaining data by special arrangement.

Make checks and money orders payable to COMMERCE/NOAA/NGDC. Do not send cash. All non-U.S.A. orders must be in U.S. Dollars drawn on a U.S.A. bank. A ten-dollar (\$10) handling fee is required for delivery outside the U.S.A. Orders may be charged to American Express, MasterCard, or VISA by telephone, letter, or fax. Please include a credit card account number, expiration date, telephone number, and your signature with order. Address inquiries and orders to:

National Geophysical Data Center
NOAA, E/GC1, Dept 891
325 Broadway
Boulder, CO 80303, U.S.A.
Phone: 303-497-6900

Fax: 303-497-6513

Internet: info@mail.ngdc.noaa.gov

Telex: 592811 NOAA MASC BDR

For questions about Global Ecosystems Data on CD-ROM please contact: John Kineman, 303-497-6900, jjk@mail.ngdc.noaa.gov

THE GEOWEB PROJECT

The GeoWeb project <<http://wings.buffalo.edu/geoweb/general.html>> was started in May 1994 and is in the developmental stage. It is being spearheaded by Brandon Plewe of the State University of New York at Buffalo as his project for an M.A. in Geography (plewe@acsu.buffalo.edu) and Chris Stuber of the U.S. Census Bureau (Chris.Stuber@census.gov) as part of his work in building better interfaces to census data. The mission of GeoWeb is to provide a service which geographers, GIS users, and the general public can use to locate geographic information that has been made available on the Internet. To fulfill this mission, GeoWeb consists of two parts: an index of Internet-accessible geographic information and one or more WWW interfaces to this index.

Internet resources in the index will include those for the general public (e.g., general information about countries, states, and places; simple maps of areas; lists and maps of Internet resources in an area) and those for cartographers and geographers (e.g., cartographic/GIS base map files, thematic data of a geographic nature, and GIS data sets).

This index will contain metadata (based on FGDC standards), giving a brief description of each data set, and a URL of its location on the net. The plan is to have the index be at the data set (not the series) level; for instance, each 1degx1deg DEM would have a separate entry (several hundred for the entire series). A separate index of more complete metadata for an entire series or URLs to series metadata pages on the WWW could be included in the future. The first phase will only include data for the United States, but the goal is to make it scalable for worldwide metadata.

The database will probably be based on the emerging Spatial WAIS search engine. However, it will need useful gateways to make it accessible to the public. Four possibilities have been discussed:

1. *Direct WAIS search.* A form with several fields would be presented (including spatial), which the user could fill out and submit. A page would be returned listing the resources which matched the criteria with hypertext links for each entry to more complete metadata and the resource itself.
2. *Dynamic map window.* A WWW page displays a simple map which can be scrolled and zoomed (a la the Xerox Map Server). At any time, the user could select a link to see a list of the resources that cover the displayed region.
3. *Gazetteer.* Based on a U.S. Gazetteer. The user would do a name search for a place (state, county, city, etc.), receiving in return a brief description of the place (from GNIS or census data) and a list of the resources that cover that place.

4. *Hierarchical Directory*. Based on either a menu or map format (similar to the Virtual Tourist). The user would first get a directory of states (or countries). Selecting one would bring up a new page containing links to a list of state-wide resources and the counties in the state. Selecting a county would display a page of county-wide resources and places in the county. Selecting a place would display a list of resources covering that place.

A paper by Brandon Plewe (1994) on the GeoWeb Project can be accessed at <http://wings.buffalo.edu/~plewe/paperwww.html>

TMS EXPERIMENTAL BROWSER

Brandon Plewe has also created a program called the TigerWWW for the Census Bureau's TMS Experimental Browser <http://tiger.census.gov/>. This site is the Washington, DC area from TIGER/92 data. Other preset values are The Mall in Washington, DC, the United States, Northeast U.S., and New York City. The interface prompts for a point location (a latitude and longitude) and a width and height in degrees for the map. Zooming in and out, specifying a point, and placing markers are all supported for the displayed map. This is an example of access to a particular set of cartographic data by the spatial description of a point and an area around that point. The approach shows promise.

NOTES

- ¹ This group is also known as the Interagency Working Group on Data Management for Global Change (IWGDMGC).
- ² For more information about the activities of the Library Information Subgroup, send email to gcdis.lis@earth.usgcrp.gov.
- ³ GRASS (Geographical Resources Support System) is public domain software developed by the U.S. Army Construction Engineering Research Laboratory (USA-CERL) in Champaign, Illinois.
- ⁴ GeoRef is a database created by the American Geological Institute that covers the worldwide literature on geology and geosciences. It is available through various database vendors and on CD-ROM.
- ⁵ An interesting project along this line is the NASA-funded project entitled "Compression and Progressive Transmission of Digital Images." The principal investigator is Dr. Jeffrey W. Percival of the University of Wisconsin-Madison who can be reached at jwp@bernie.sal.wisc.edu.
- ⁶ Semantic net technology uses dictionary definitions and thesaurus links to build relationships—both word-form and meaning-based relationships—between words so that the words of a query can be expanded into related terminology for the search of the database. ConQuest Software is located at 9700 Patuxent Woods Drive, Suite 140, Columbia, MD 21046.
- ⁷ The address of E-Systems Inc. is Garland Division, PO Box 660023, Dallas, TX 75266-0023.

REFERENCES

- Ben-Avraham, A.; Kempler, D.; & Ginzburg, A. (1988). Plate convergence in the Cyprean Arc. *Tectonophysics*, 146, 231-240.
- Committee on the Environment and Natural Resources (draft of July 7, 1994). *The U.S. Global Change Data and Information System Implementation Plan*.
- Hill, L. L. (1990). *Access to geographic concepts in online bibliographic files: Effectiveness of current practices and the potential of a graphic interface*. Unpublished dissertation, School of Library and Information Science, University of Pittsburgh, December.
- Nebert, D. D. (1995). *Use of Z39.50 to search and retrieve geospatial data*. Manuscript submitted for publication. Preprint available from <http://130.11.51.187/public/DLIpaper395.html>
- Plewe, B. (1994). *The GeoWeb project: Using WAIS and the World Wide Web to aid location of distributed data sets*. Available from <http://wings.buffalo.edu/~plewe/paperwww.html>
- Woodruff, A. G., & Plaunt, C. (1994). GIPSY: Automated geographic indexing to text documents. *Journal of the American Society for Information Science*, 45(9), 645-655.

Geospatial Information Needs of the General Public: Text, Maps, and Users' Tasks

Geospatial data are pervasive and the methods of understanding how people resolve their spatial information needs have been diverse. This discussion reviews the role of spatial information in resolving information needs. We add to this knowledge base the results from a series of ongoing experiments. These experiments address how geospatial information is used and how it supports resolution of the information needs of the general public. Five exploratory experiments are briefly reported revealing significant relations among format, task, and situation in geographic contexts. Based upon these preliminary outcomes, an extended role for libraries to add value to their services and products for geospatial information is proposed. Also, suggestions are made for future directions for research in this critical domain of geospatial information seeking.

INTRODUCTION

Geospatial information is very pervasive. As a minimum, we seek weather information and driving or walking directions as well as airline and travel information. Often we do not even think of the information as geospatial in nature: What are those campsites like in the Florida Keys? What time can I arrive in Champaign for that conference? And many more.

This discussion integrates and summarizes the findings from five studies I have conducted over the past few years. Each of these exploratory experiments addressed different aspects of four general research questions:

1. What are the geospatial information needs of the general public?
2. What are the different formats and tasks for geospatial information suggested by the public?
3. What formats for geospatial information are most useful under differing task situations?
4. What role may the public library play in assisting to resolve geospatial information needs for the public?

The results from these experiments are not the definitive answers to these questions. They clarify the questions and provide insights to the phenomenon framed by these questions. A negative or deconstructionist

assumption guides this research program: Public libraries have not met the geospatial information needs of their patrons as well as they might. This project also claims that the geospatial information needs of the public have not been systematically analyzed. Public libraries have determined which books, magazines, videos, CD-ROMs, and paper maps they should acquire. This is in contrast with libraries looking at their patrons' geospatial information needs and developing a collaborative role with patrons in resolving needs with broad-based processes, products, and services. This work further assumes that deeper understanding of patrons' geospatial information needs will permit public libraries to better assist their patrons in resolving these needs.

BACKGROUND AND MOTIVATION FOR THE STUDIES

Definition of Geospatial Data, Information, and GIS

Geospatial data are data that describe the earth at a scale from the size of a room to the whole planet. Such earth data may be collected from field studies, aerial or satellite photographs, or remotely sensed. *Geospatial information* is the process of converting geospatial data into a form that helps users to resolve their particular information needs. The user may do some or all of the data-to-information conversion, or information providers may add value to the basic data. However, the users must be the ultimate determiners of the relevance of the data and information to meet their needs.

For centuries, geospatial data and information have resided in libraries as either paper maps or in text. Often these materials were amassed into atlases, gazetteers, or other collected works. More recently, maps and geospatial text have been stored in microforms that retain most of their historical functionality while taking up less storage space. Much of the latest geospatial information is being created and stored in digital form as text and map graphics and in geographic information systems (GIS).

One definition of a GIS is a computer system designed to allow users to collect, manage, analyze, and display large volumes of spatial reference data and associated attribute data (Guptill, 1988). A GIS employs tabular data about spatial and nonspatial entities as well as the topological data that indicate how those places are connected to each other. Spatial entities may be places corresponding to points (e.g., location of an oil well, crossroads), lines (e.g., streets, railroads, airline routes, streams), or polygons (e.g., lakes, corn fields, range for an owl species, university campus). Scale is important in this regard. For example, a city may appear as a point on a one page map of the United States or it may appear

as a polygon on a one page map of a county. GIS further provides tools to manipulate, analyze, and display attribute and topological data thereby producing geospatial information.

Maps and Map User Tasks

Maps, whether paper or digitally displayed, provide powerful representations of the real world. However, these are always misleading because of the conversion from a three-dimensional spheroid to a flat surface (Monmonier, 1991). People's perceptions of their own environment are also distorted by their experience and attitudes. In 1960, Lynch, in his book *The Image of the City*, described an experiment he conducted to record how urban residents viewed their city. Residents' perceptions differed depending upon where in the city they lived. For example, those who lived in more affluent parts often omitted or radically distorted the size and shape of nearby high crime areas. Lynch, in later work, also showed that what people pointed out when walking was not necessarily what they commented upon when traveling by car (Appleyard et al., 1964).

We humans develop spatial understandings of our surroundings by two major methods: procedural and survey. Some of our knowledge of the environment derives from route knowledge gained by traveling around the area, learning what is connected to what. Such connections may be different for different people depending upon the routes traversed and the frequency of traversal. Such learning by "feel" is procedural knowledge. We also acquire knowledge of spatial relationships by looking at maps and getting survey or overview knowledge. The merging of survey and procedural knowledge of the environment is not immediate for most people. For example, people frequently know the turns to take but not the name of the streets when giving directions even in a town they have lived in for quite some time. Also, people will often know the distance between two landmarks but not the route between them. Also, many of us, for a variety of reasons, take different driving routes to work than we take home from work (Vaziri & Lan, 1983).

The tasks we wish to accomplish influence our knowledge of space. Thorndyke and Stasz (1980) present the results from one of their experiments in which subjects were asked to memorize the contents and relations in a map and were tested on their recall. One subject did particularly poorly on the recall of names of places. During debriefing, he stated he was a military pilot and from flying he learned to recognize places by visual inspection and not by their names.

Also, text and graphics influence our perceptions of a place. Few of us would plan expensive vacations without seeing photos or a video of the area we wished to visit or without listening to the current news from

the area. Sarajevo was once a beautiful city; Lebanon was once a glorious Middle Eastern resort paradise. What are your feelings about visiting Miami, Florida? Would you rather go in winter or summer? Is the land you are about to buy in a flood zone? Do you wish to move to California near the San Andreas Fault line? Where we visit, where we live, and how we make decisions in our lives are constantly affected by geospatial images and text.

The methods for assessing the effects of various formats of information are not well understood. Larkin and Simon (1987) introduced the notion of *information equivalence*. A graphic and text are deemed information equivalent if the same tasks may be accomplished using either format. Larkin's work explored performance in the use of force diagrams compared to word descriptions to solve elementary physics problems. Information equivalence is not the only measure of the effectiveness of information provision from different formats. Foreign language translators frequently assess the comparability of a translation with the original text using the measure "informativeness" (Lehrberger & Bourbeau, 1988). Informativeness assesses the quality of the translation and its faithfulness to the ideas in the original.

Many of the experiments using maps or geospatial information have been guided by human information processing models (Lindsay & Norman, 1977) and only assess the time on task and accuracy of map use (e.g., Lloyd & Steinke, 1986). In these models, human cognition is represented as analogous to a computer with inputs, processing, and outputs. These models rarely seek the strategies subjects employed during manipulation or the deductions subjects perceived during and after the tasks. Also, most studies that do collect users' impressions frequently only do so during a debriefing after an experiment (e.g., Thorndyke & Hayes-Roth, 1982; Thorndyke & Stasz, 1980). Those studies that do formally collect users' impressions, deductions, inferences, and understandings often ask what users' might do, not what they actually do or have done (e.g., Vanetti & Allen, 1988; Gould & White, 1986). Different formats also affect human performance and the selection of strategies to accomplish information retrieval tasks. For example, Benbasat and Dexter (1986) have shown that people extract more precise and accurate values when they read from a table of data than when they use a bar or line chart. They also found that the use of charts produces more rapid data retrieval, and people are better able to detect trends in data with charts than when they use tables of data. In another study, Bieger and Glock (1986) found that subjects who used a diagram to put simple toys together were faster but more error prone than people who used text instructions for the same assembly process. In an experiment involving navigating by car, Streeter, Vitello, and Wonsiewicz (1985) found that subjects who were given audio navigation instructions (when to turn and distance to next turn) committed fewer errors and found their destinations more rapidly than did subjects who used traditional road maps or enhanced stylized route maps.

Spatial information needs and tasks are of several kinds. There have been several simple classifications proposed as well as more detailed schemes. One simple approach states that geospatial information involves only finding responses to two questions: What's there? and Where's that? These include finding what is present at a particular location and all locations of a particular phenomenon.

Muehrcke (1986) suggests another approach to the categorization of geospatial information tasks. He posits that geospatial queries may be matched to the levels of educational objective taxonomies (e.g., Bloom, 1956). Muehrcke describes three levels of sophistication of geospatial queries: reading, analysis, and interpretation. Reading involves extracting facts from a text or map such as the distance between two points, the name of a place, or latitude and longitude of a place. Questions about reading generally answer "what" questions or have "yes or no" responses. Analysis involves looking for relationships or seeing patterns in text or maps, such as noting that people living near polluted water supplies have high incidents of intestinal bacterial infections, or establishing a pattern for fossils distributed among rock layers. Analysis questions generally answer most "how" and relationship questions. Interpretation questions seek cause and effect such as noting that cholera is transmitted from individual to individual by water or that all these trees were damaged because of the explosion of Mt. St. Helens. Interpretation questions answer "why" and some "how" questions.

Several schemes have been developed to understand map design variables better and improve the ability of maps to help users with their information needs (e.g., Bertin, 1981; Armstrong et al., 1992). These schemes investigate how different variables such as size, shape, color, texture, gray scale, orientation, and position affect the usability and usefulness of maps.

The studies briefly described earlier emphasize the importance and usefulness of spatial information whether in text or maps, electronic or paper. They also clearly show that we still do not understand how best to organize or present geospatial information for a particular user with a particular information need. Additional studies that extend our current knowledge and provide new insights are needed. Toward that goal, we have begun to conduct a series of exploratory studies addressing different aspects of the role of maps, text, and user tasks in relation to geospatial information needs. We report the findings from five of these ongoing studies that we feel are especially significant at this time for library and information professionals.

A SERIES OF EXPLORATORY STUDIES

Map-to-Text and Text-to-Map Translation Exercises

Translation Method. To explore the degree to which maps and geospatial texts may be informationally equivalent, we conducted an

exploratory experiment. We assessed the degree of informativeness among maps and geospatial text as perceived by a small group of subjects with varying geographic backgrounds. To assess the equivalence of maps and text, subjects were given a set of four geospatial materials: two paper maps and two texts with geospatial content. The four geospatial materials consisted of weather information, a description of a local hiking trail, a description of rural vehicle flow rates, and the layout of a small town (reproductions of the materials are available from the author).

Two sets of subjects participated in these exercises. Eleven subjects participated in total, six in the first set and five in the second. Eight of the subjects were students in a school of information studies with no special training in geography or map reading, and three subjects were geography graduate students. These subjects were a small sample of adults with a range of expertise in geospatial information.

The first set of subjects drew maps from the weather and hiking trail text descriptions and wrote text descriptions for the town plat and the vehicle flow maps. The second set of subjects wrote descriptions for the weather and trail information while reviewing the maps, and drew maps for the plat and flow information while reviewing texts. That is, subjects translated, as best they could, the information in one format into the other. Subjects also evaluated their translations by comparing them with the originals and scoring their own products on informativeness scales. Subjects also recorded the time they took to produce each translation.

Translation Results and Discussion. Table 1 reports subjects' own evaluations and their time on the task for each translation. Subjects frequently had deep concerns about their attempts to translate from text to map or from map to text. However, none felt that the task was impossible, although their assessments of the quality of their translations were generally low. Low informativeness ratings (values near 1) suggest that subjects felt that their own effort left out much information from the original while high ratings (values near 5) suggest that subjects felt their translation contained most of the information from the original.

Major reasons subjects gave for low informativeness ratings of the map-to-text exercises centered on the number of words needed, the time it would take to complete a quality translation, and the complexity of the map. The least complex map/text pair was the town plat, and it received the highest informativeness scores, partially corroborating these reasons.

Major reasons subjects gave for low informativeness ratings of the text-to-map exercises included subjects' self-assessment that they could not draw well, their lack of recall of standard symbols and icons, the difficulty in showing temporal change, and their inability to express weather statistics without words. Several subjects protested that the texts were not complete enough to construct a high quality map. This was especially

true for the hiking trail and the traffic flow texts. Subjects also mentioned that the text of the hiking trail description lacked "organization" and was quite complex (although the hiking trail text came from a well-reviewed guide to hiking trails).

Subjects perceived no statistically significant differences in the informativeness of translations between text and map production except in the trail translation. For the trail, subjects felt they produced more informative texts than map translations (*t* test for independent samples with $\alpha=.05$; $df=9$). This difference was driven by the difficulty in producing the trail map from the text. Although not statistically significant, a similar trend is seen in the creation of the weather map from text. The only large differences in the mean times for translation were in the creation of text for the town plat and the traffic flow, although these were not statistically significant with such a small sample. These were also the two maps with the least amount of details.

TABLE 1.
INFORMATIVENESS AND TIME ON TASK FOR TRANSLATION EXERCISES

Translation & type	Informativeness				Time on task			Number of subjects
	Mean	Median	High	Low	Mean	High	Low	
					(minutes)			
Weather Text-to-Map	1.8	2	3	1	11	13	8	5
Weather Map-to-Text	3.0	3	5	1	15	20	12	6
Town Text-to-Map	3.7	3.5	5	3	16	25	9	6
Town Map-to-Text	2.8	4	4	1	8.5	10	5	5
*Trail Text-to-Map	1.2	1	2	1	17	20	12	5
*Trail Map-to-Text	3.3	4	4	1	19	28	12	6
Traffic Text-to-Map	2.6	2.5	4	1	16.5	25	12	6
Traffic Map-to-Text	3.2	4	4	1	9.5	13	5	5

*Only pair with statistically significant difference.

Focus Group Discussions of Map and Text Translations

After completing the map or text translations, the subjects participated in focus groups discussing how they manipulated the information and constructed their translations. They also discussed their personal map use.

Question #1. What are your major personal uses of geographic information?

Responses to question #1:

- recreational travel
- survey knowledge to become acquainted with an area
- weather information
- building layouts: fire exit plans, mall store locations, etc.
- sports needs: trail riding, hiking paths, ski areas

- recreational games: adventure, military strategy, etc.
- imagining/dreaming/fantasizing of places to visit

Participants also expressed the human needs that stimulate seeking out maps in subsequent question responses including:

- seeking food
- connecting with others
- seeking aesthetic surroundings, and
- residence selection

Question #2. When presented a map, how do you act to make sense of it? What do you do first? Next?

All but one participant agreed that they first look at the title and legend and then seek out the content that addresses their information needs. Such needs included way finding, survey knowledge, labels (e.g., names of streets, towns), elevation, terrain, scenery, and so forth. They inspect the base map checking relative positions and directions and then obtain metric information such as distances or compass headings. Top-down or bottom-up explorations of the maps seemed to depend upon the task. For example, most agreed that way finding was top-down and then point to point, while survey knowledge was usually found by going back and forth between the big picture and focusing on small regions or points. Participants also said that, for a given information need, several maps may often be needed: a highway map may give the fastest route but may not show all the points of interest or scenery.

Question #3. Do you feel text and maps can express the same geographic information? How do texts and maps communicate geographic information?

Participants felt geospatial text and maps do different things. This did not contradict their belief that one can be translated into the other with much hard work. Texts give inventory and specific site descriptions but not spatial relations. They agreed maps are succinct representatives useful to express spatial or topological relationships (e.g., what is next to what, what is inside what, what is between what).

Participants felt the translation processes are possible, but many words are needed to make the text informationally equivalent to a map. Also, the more abstract the ideas the harder they are to draw. They also indicated that the display of temporal information in a map was especially difficult (for example, showing temperature changes over time on weather maps). They did, however, feel that these difficulties could be overcome with multiple maps or map animations.

Some participants felt that text is more human and more imaginative. For example, "go north" seemed less human than "go to the Pan Am building." Most felt a mix of map, video, and text is best for expressing geographic information. For example, participants said using only map symbols made it hard to express a sixth record breaking high temperature yet it would have been simple in words. Participants felt translating a map to text was easier than translating a text to map. They agreed that this may be because all had done more writing than map drawing in their lives.

Question #4. When do you put a map away?

The trivial answer was when they arrive at their destination. However, all participants agreed they often use maps iteratively. Due to information overload and short term memory limitations, they get some information from the map, put the map down to think or act on that piece of information, and then pick up the map again for additional information or details.

In summary, participants felt that most information in text or map may readily convert to the other. However, some information in each is very difficult, if not impossible, to convert to the other. Consequently, from the point of view of this very small sample, geospatial text and maps are not informationally equivalent. Even if they were, some tasks are much better served by one format than the other. Maps serve to support many tasks but not an infinite number, and the tasks appear to be categorized and manageable.

Geospatial Sense-Making Situations of Patrons

Talking to people about their actual geospatial information need situations provides interesting insights for library processes, products, and services. This section reports partial results from a study in which people sought geospatial information and described their situation, the events, questions, tasks, and formats of information they found and used. A sense-making time line method and survey instrument were employed to collect this data.

Sense-Making Method and the Sample. Sense-making is a structured open interview and survey technique that captures the experiences of people in their own words (Dervin, 1983, 1992; Dervin & Nilan, 1986). Sense-making time lines provide a sequence of the events that occurred in resolving an information need in a particular context or situation. The method addresses the overall process of the situation claiming that situation is the most important determinant of information need. Therefore,

the analysis is not dependent upon the total recall of any one individual since many people find themselves in similar situations.

Dervin's approach explicitly accounts for time and space dependencies of human information needs and processing by metaphorically representing the cognitive states of humans as movement along a road in time and space. The perception of a person's current position along this road is a function of where a person has been (experiences, environment, and so forth), where the person is (present), and where the person is going (future). Dervin pictures humans as intelligent creative creatures capable of making sense by incorporating knowledge from within themselves and from the external world allowing for forward movement along their cognitive road. She notes that gaps often appear along the cognitive road that represents a need for the individual to make sense of the world before movement along the road may continue. Gaps are a direct consequence of Dervin's perspective of a human's view of reality as intersubjective and constantly changing. These gaps are user concerns that generate an information need with users' expectations of how bridging the gap would be helpful.

The present research assumed that geospatial information needs as seen by the user are most easily discovered by focusing on the gaps in users' cognitive roads. Dervin's approach allows each person to represent his or her own reality; however, Nilan (1985; Newby et al., 1991) suggests that this does not lead to chaos in analyzing human behavior because humans are connected by shared sociocultural experiences. This overlap of experience limits the range of diversity present at a particular point in time. The time line and sense-making metaphor focuses on points in time where information is needed. Thus, scientific prediction about human behavior changes from trying to make "If ... Then ..." generalizations to predicting a sequence of events in the form "Then ... Then ...": first one does this and may need this, then one does that and may need that. This results in developing a sequence that gives users what they need when they need it (Nilan et al., 1989).

Operationalization of Dervin's model for this experiment involved respondents describing the time line of events they experienced while resolving a recent need for geographic information (the situation). Events along the time line may be viewed as frames from a filmstrip of the situation, and the gaps along the cognitive road are the manifest questions the users raised at each event.

Eighty-two subjects completed a short sense-making time line survey describing a recent use they had for geospatial information. The subjects varied demographically, educationally, professionally, and, significant for this work, in their geographic education and geographic professional background (Gluck, 1993, 1995). Subjects' ages ranged from 14 to 59 with a

mean age of 25 and median age of 21; one-fourth were above 28 and another one-fourth below 21. Several were in high school, several were employed in positions that involved geospatial information such as travel agent, delivery person, or professional geographer. Only one-fourth of the subjects had such professional geographic experience or formal geographic education beyond high school. Forty-three males and thirty-nine females volunteered to participate in the group sessions using the self-administered sense-making survey.

Results of Geospatial Sense-Making. A detailed content analytic scheme was developed for the eighty-two situations described by the respondents in their sense-making time lines. The scheme employed the situation descriptions as stated by the respondents. A set of twenty-five (30 percent) of the descriptions composed a random sample for development of the scheme. The scheme was then refined with scope notes and intracoder testing was done (with more than 90 percent intracoder reliability) and applied to the remainder of the situation descriptions. For reporting purposes, the broad scheme is condensed to four major categories and the major subcategories are described when meaningful.

The four major categories were:

1. *Educational Geospatial Information Need Situations.* This category included data needed for class assignments, spatial tool training, pilot training, and internship needs.
Example: "Needed to find geographic information to make a map of Vermont for an assignment."
2. *Professional/Career Geospatial Information Needs Situations.* This category included attendance at conferences, career planning, job interviews and relocation, delivery services wayfinding, field data collection as part of employment, and seeking businesses and business location analysis as part of employment.
Example: "I was driving from Ridgefield, CT to Hartford, CT (approx. 1 hour) for a job interview."
3. *Personal Geospatial Information Needs Situations.* This category included emergency needs (e.g., vehicle breakdowns), moves to a new community, travel and trip planning to visit family or friends or to communicate such information to others, general travel planning and execution, and seeking weather information for planning purposes or mid-course corrections while traveling. This category did not include vacation, recreational, or career related situations, rather it emphasized connecting with family and friends.
Example: "I had to visit my friend at his school, and I had to find my way."
4. *Recreational Geospatial Information Needs Situations.* This category included vacation, honeymoon, and mixed business/pleasure travel planning and execution, finding a place to go when bored,

fantasizing/imagining/daydreaming about other places, travel with children, and weather information for sports activities or to get out of the elements. Example: *"We took a day trip for wine tasting in the Finger Lakes."*

Of the eighty-two situations, seven (8.5 percent) were educational, sixteen (19.5 percent) were professional/career, twenty-two (26.8 percent) were personal situations, and thirty-seven (45.1 percent) were recreational with most of those twenty-three (28 percent) execution of vacation travel. Of these situations, only seven (8.5 percent) of the responses involved the use of libraries. Those that did involve libraries explicitly sought geographic data for class assignments (four) or employment information (three). Others may have used a library but did not say so explicitly and did not find the responses to their questions in a library.

As a first level of task understanding, Muehrcke's three category scheme of reading, analyzing, and interpreting geospatial tasks was used to describe the respondents' questions. We only applied this analysis to the questions within the two most important events of their information seeking situation. This emphasis permitted us to concentrate on the respondents' most important information concerns.

The eighty-two respondents recalled 185 questions from the two most significant events in their situations, an average of 1.13 questions per event. This is lower than found in face-to-face sense-making interviews (e.g., Schamber, 1991; Newby et al., 1991), and represents the tradeoff in using a self-administered protocol. The protocol was easier to administer than an interview process but yielded less in-depth responses; however, the range of tasks, formats, and events described in these eighty-two situations is sufficiently rich for thorough analysis. Applying the reading, analysis, or interpretation task category scheme was quite straightforward. Therefore, only one coder was required to complete assignment of values. The intracoder reliability was greater than 90 percent. For example:

Reading questions included:

- Should I continue or return?
- What is it going to cost to get there?

Analysis questions included:

- What kinds of rocks do these fossils hide in?
- How could I ascertain spatial resolution fine enough for this task?
- Where could we find a restaurant?

Interpretation questions included:

- The travel agent is aware of time differences that caused the luggage foul-up. Why can't they prepare for this?
[Respondent lost luggage on his honeymoon.]

GLUCK/TEXT, MAPS, AND USERS' TASKS

- Why are there so many dead-end streets in Jersey City?
[Respondent was lost in Jersey city trying to return to Statue of Liberty pier.]
- How could so many hardships be brought upon the Jewish people?
[High schooler visiting a museum in Israel.]

Of the 185 questions, 113 (61 percent) involved reading tasks, 47 (25 percent) involved analysis tasks, and 25 (14 percent) involved interpretation tasks. Over one-third of the questions involved more than seeking factual information that could be directly read from a map or geospatial textual description.

Respondents reported the format in which they found each response as part of the sense-making protocol. The format categories for analysis were developed using an inductive content analysis scheme applied to those responses. Table 2 reports the format categories as seen by respondents and their frequency of occurrence.

TABLE 2
FORMATS OF RESPONSES TO GEOSPATIAL QUERIES FROM SENSE-MAKING (N = 185)

<i>Format category</i>	<i>Count</i>	<i>Percentage</i>	<i>Cumulative Percentage</i>
Used map alone	19	10%	10%
Observed in the environment	13	7	17
Thinking, resolved it myself	19	10	28
Used written text alone	17	9	37
Spoke to someone else	69	37	24
Combinations of others	17	19	83
No response found at time	31	17	100

Many more questions were answered by speaking with others than any other format of response (69 of the 185 or 37 percent). Respondents sought out other people to answer their questions rather than use other sources and formats. Maps used alone were important but only accounted for nineteen or 10 percent of formats in which respondents found responses. Mixed formats that may also have included maps accounted for barely another 9 percent of responses. Thinking about it (self) accounted for another 10 percent while observing it in the environment accounted for another 7 percent or 13 responses. Respondents did not find responses to their questions at the time they had them for 31 of the 185 questions (17 percent).

Respondents made satisfaction judgments for the degree to which both the responses and the process of obtaining them satisfied them. Overall, no significant differences in satisfaction were found controlling for tasks or formats; that is, respondents were just as satisfied with the responses they found

whether they were doing reading, analysis, or interpretation tasks, or whether the responses they found were from maps or texts.

Relevance judgments were also made for the responses obtained. However, the relevance codes were assigned by the researchers who developed the other content analysis schemes (Gluck, 1993, 1995). Some relationships found among relevance judgments, formats, and tasks included:

- Maps provide significantly more relevant information than text, especially in analysis tasks.
- Responses to text analysis are judged less relevant than map reading or analysis responses.
- Text does provide highly relevant responses especially for geographic reading and interpretation tasks.

Competence in Geospatial Information with Tasks and Formats

Method of Testing. In another of our exploratory experiments, subjects performed reading, analysis, and interpretation tasks with geospatial information similar to a classroom learning experience. The same eighty-two subjects who provided geospatial sense-making data volunteered to participate in a paper and pencil test of their abilities. Data on subjects' competence with maps were compared with their competence with text on covering a range of geospatial reading, analysis, and interpretation tasks.

Subjects read or heard a text of geospatial information and viewed maps representing the geography of the text information. Subjects reviewed two sets of materials, one set described the spread of cholera in South America in the early 1990s while the other set described the rise of banking in Florence, Italy, in the thirteenth century. Half the subjects heard the texts and the other half read them. Both sets of subjects reviewed the same paper maps. Reading, analysis, and interpretation tasks involved answering short answer and/or fill-in questions based on the contents of the maps and texts. Information to answer the questions and perform the tasks appeared either in the maps or in text but not in both. Accuracy of responses and the time to complete each task provided a basis for competency scores for each subject.

Competency scores were the ratios of the number of items correct over time to complete the tasks, similar to the words per minute competency of a typist. Basic competencies were formed by using all combinations of two formats (maps or text) with three tasks (reading, analysis, and interpretation) with two forms of text presentation (written or oral), resulting in twelve measures (2x3x2). Preliminary analysis showed that oral or written presentation of the materials did not affect the mean time

or accuracy of subjects' performances (t test; $df=80$; $\alpha=.05$). Basic competencies were then combined to form task, format, and overall competency scores.

Competence Testing Controlling for Format and Task. We employed repeated measures within subjects' analysis of variance statistical procedure to analyze the competency scores. The analysis also controlled for task and format (Gluck, 1993, 1995). A significant main effect was found for task: subjects were significantly more competent in reading tasks than in analysis or interpretation tasks, and significantly more competent in analysis tasks than in interpretation tasks on average ($\alpha=.1$; $N=492$). Each of these significant differences of means had effect sizes above one-half of a standard deviation, showing an important difference (see Levin & Lesgold, 1978; Levie & Lentz, 1982). Format had no main effect; that is, subjects were as competent with maps as they were with texts on average. However, subjects did significantly better with some format task combinations than others. Subjects were most competent on text reading, less competent on map reading, map analysis, and text analysis tasks, and were least competent on both map and text interpretation tasks. These differences were also in the range of at least one-half of a standard deviation showing an important difference. This analysis was also quite reasonable since it accounted for 46 percent of the variance in the performances (R -squared = .46).

Experience of subjects in personal, educational, and professional use of geospatial data was also collected. The relationship between experience and competence controlling for format and task was complex. Our analysis shows experience does affect competence, especially in reading and analysis tasks. The analysis also shows that formal education alone does not support subjects' ability to perform effective interpretation tasks compared with their overall life experiences (Gluck, 1995).

Log of Map Use in a Public Library

What sort of maps do public library patrons request when a wide range of local paper maps are made available? The Leon County Public Library in Tallahassee, Florida, with support from other county agencies, including the Geography Department and other offices at Florida State University (FSU) created a Map Resource Center (MRC) in 1994. Volunteer interns from the Department of Geography at FSU staffed the MRC afternoons, evenings, and weekends. Also, the MRC is open for self-service during regular library hours of operation. This library serves Florida's state capitol and surrounding rural areas of Leon and adjacent counties. The population served exceeds 150,000. The staff of the MRC maintained a log of maps used and some descriptions of their interactions

with patrons. The log from this one library is not very generalizable; however, a review of its contents points to issues and concerns that may be true for a wide range of public libraries with diverse and multicultural constituencies.

The paper maps used and general questions asked of the MRC personnel were recorded for the period June 1994 through February 1995. The record of activity at the MRC was an informal log that was not designed for research purposes. The log was only an informal record of activity, and our analysis of the log is a post hoc content analysis.

Table 3 displays a list of categories we found to describe the materials selected by MRC staff. The North Central region of Florida had severe flooding and rainwater runoff problems during the period the log was maintained; consequently, most of the uses of the environmentally sensitive area (ESA), contour and FEMA maps related to the natural disaster, such as patrons trying to decide if they would flood or verify flood insurance rates. Contour isoline maps showed two foot differences in elevation. Also, frequently ESA, flood, and FEMA maps would be consulted jointly to conduct an ESA check for a patron. The purpose for the ESA check was rarely explicit in the log. We considered a general tourist map as one that did not involve an explicit reference to roads or street maps in the log. The other geological, topographic category included only general reference in the log to geology, elevation, or topography. We also listed the explicit use of USGS quads as a separate category. Zoning maps were those that reflected local zoning ordinances. Recreational area map use included seeking information on trails for hiking or biking, on fishing sites in recreational or national forest areas, and general information about these areas. Naturally, some patrons were shown more than one product in a visit while others returned to look at materials more than once. Many entries in the log confirm the suggestions from the focus group participants mentioned earlier (the section on Focus Group Discussions of Maps and Text Translations) that multiple maps are frequently needed to resolve a spatial information need.

Table 4 displays a scheme that categorizes patrons' explicit purposes for map use. The comments listed in the log by MRC staff and used by them for map selection permitted development of this category scheme. Additional purposes may have been expressed to the MRC support staff but not recorded in the log, and many purposes may not have been expressed by the patrons. Approximately one-third of the entries had determinable patron purposes. If no purpose was explicit in the log, none was included in Table 4. Twenty-eight of the entries in the log indicated that the library did not have the requested item. The log also contained twenty-five explicit entries referring patrons to other agencies, such as the state library, for historical maps or to the county appraiser for parcel ownership information.

GLUCK/TEXT, MAPS, AND USERS' TASKS

TABLE 3
MRC FREQUENCY OF USE BY CATEGORY (JUNE 1994 THROUGH FEBRUARY 1995)

<i>Paper cartographic products consulted</i>	<i>Local</i>	<i>Florida</i>	<i>U.S.</i>	<i>Other</i>	<i>Total</i>
Appraisal/parcel	21	0	0	0	21
Astronomical charts	3	0	0	0	3
Environmentally sensitive area (ESA)/ Contours/FEMA	99	1	0	0	100
General tourist	5	6	11	17	39
Historical	3	4	3	5	15
Local comprehensive development plan	2	0	0	0	2
Other geological, topographic	9	3	4	3	19
Political district boundaries	0	1	0	0	1
Recreational areas	15	1	3	0	19
School district boundaries	10	0	0	0	10
Soils	4	0	0	0	4
Street/road	23	24	20	1	68
USGS Quadrangles (7.5 min)	33	7	1	0	41
Zoning	17	0	0	0	17
TOTALS	244	47	42	26	359

TABLE 4
MRC LOG EXPLICIT PURPOSES FROM PATRON FOR MAP USE

<i>Patron purposes explicit in MRC log</i>	<i>Local</i>	<i>Florida</i>	<i>U.S.</i>	<i>Other</i>	<i>Total</i>
Genealogical investigation	4	0	0	0	4
Mileage computation	1	1	0	0	2
Place finding	5	0	5	0	10
Professional/career	3	0	0	0	3
Radon check	1	0	0	0	1
Real estate purchase	12	0	2	0	14
School assignment	10	10	11	35	66
Trip planning (ground/air)	1	0	4	1	6
TOTALS	37	11	22	36	106

When assistance with analyses or interpretations were available, patrons sought them. Examples here include environmentally sensitive area (ESA) checks and flooding patterns. Almost one-third of the purposes were analyses for real property while another large category was school assignments. Reading tasks were rarely explicit in the log; however, the number of street maps indicated in Table 3 imply many reading tasks motivated patrons to seek out geospatial information.

GENERAL DISCUSSION OF THE EXPERIMENTS AND IMPLICATIONS FOR PUBLIC LIBRARIES

We now return to the four questions that motivate this line of research. Unfortunately, we are unable to provide definitive answers to any

of these questions now. We can express what we know and what we need to learn including the range of methods needed to address different aspects of these phenomena.

1. *What are the geospatial information needs of the public?* Geospatial information needs vary as shown in the tasks and situations suggested in the earlier section on "Geospatial Sense-Making Situations of Patrons." However, what is most important to observe is that the number and type of categories of use and need are manageable. We do not have a comprehensive list of situations from just these two exploratory experiments, but any such list would be far from idiosyncratic and chaotic. The four major categories of geospatial information needs (educational, professional/career, personal, and recreational) with the format and task issues discussed later provide us a wide view. They also suggest how we might better service our patrons' geospatial information needs.

The affective and metaphysical needs of people appear in these categories; library services have tended to ignore these traditionally. Another need for geospatial information echoing through these studies is the need to connect with others and to imagine and dream. All geospatial information is not as teleologically or practically based as our current library services and products seem to dictate. People use the current products and services but modify them for their needs. We should be sensitive to those transformations and provide support for them.

2. *What are the different formats and tasks for geospatial information suggested by the public?* More detailed levels require further exploration, but the purposes of the earlier section on "Log of Map Use in a Public Library" and the general categories of reading, analysis, and interpretation form a very high level scheme. Just less than 60 percent of needs pose reading tasks, while just less than 30 percent pose analysis, pattern, or relationship tasks, and about 15 percent of needs pose interpretation or cause and effect tasks. We have strong indications that if more analysis and interpretation tools and materials were available, the library would be an even more welcome place to resolve geospatial information needs. As an example, consider the ESA and flooding pattern checks that brought many people into the library seeking analysis materials. When the library addressed real world needs, the public flocked to the library.

The public expresses several additional categories for geospatial information formats than our traditional library categories of text and map. They add the understanding that much information, although of a geospatial nature, comes from audio sources. The need to connect with others implies both visual connection and aural interaction.

3. *What formats for geospatial information are most useful under differing task situations?* The public does not seem to have a real concern that the answer must or should be in a particular format. They want the information in a format that is usable with the least amount of transformation necessary to get their needs resolved. Subjects readily concede that maps seem to convey information differently than text, although they find it difficult to express the distinction for all situations. Subjects expressed the power of combining text and maps for most effective resolution for a geospatial information need. Maps seem to provide much survey knowledge and to expose spatial patterns while text supports answering factual questions and clarifying cause and effect. Maps may also be more relevant for certain tasks (e.g., analysis), but they are not seen as more satisfying.
4. *What role may the public library play in assisting to resolve geospatial information needs for the public?* Beyond their traditional role in providing sources for "facts," these experiments suggest that libraries can provide added-value services and products for geospatial information analysis and interpretation. Such added-value services and products are appreciated by patrons: ESA and flood pattern checks received high satisfaction and appreciation marks from patrons. The ESA checks and flooding pattern are not hard to learn and providing short "how to" sheets may permit patrons to conduct analyses for themselves.

In addition, we can construct local support for connecting and dreaming by adding value to what patrons might wish to share with others. For example, could we not record walking or driving tour tapes of local sites for check out with the necessary equipment for their use? The list of sites could be generated by a small local survey while the cost of production might be obtained from the local chamber of commerce.

The programming to support the adding of value to current base products and extending the base of products is needed and pointed to by patron habits. Much spatial information is developed within local government, and the sharing of that information with the public is a key role that libraries can play in supporting our democracy, individual patrons, and ensuring their own viability. New formats may support these new services well. Much value may be added to local data in electronic formats such as locally developed CD-ROMs of ESA data or loading a bitmap image of the local comprehensive development plan onto Free-Nets or other local electronic providers.

SUGGESTIONS FOR FUTURE WORK

This series of exploratory experiments suggests that the current formats, content, sources, services, and structure of geospatial information

are not overwhelmingly used, usable, useful, relevant, or satisfying to public library patrons. Current formats and sources get a C+ or perhaps a B-grade if we use relevance and satisfaction measures, and perhaps an A- for assistance to resolve text-reading tasks and a B+ for assistance to resolve map-reading tasks. However, low grades must be assigned for map and text analysis task resolution support and probably failing grades for map and text interpretation task support.

Is it the patrons who fail or is it the structure and content of sources, services, and formats that fail? I would claim, from a deconstructionist perspective, that the patrons do not fail. Our historical solution to the inadequacies of spatial information usability and usefulness has been to train the user. Yet with a range of educational backgrounds, the general population does not perform most analysis and interpretation tasks well, nor do they seek much of that information in libraries. Might it be that they do not expect to find such support for higher level tasks in libraries?

Our sources and services support reading tasks moderately well but do not support analysis or interpretation tasks very well. An information driven society and an information intensive economy demand more of their citizens than to read facts. They demand that citizens perform analyses of patterns and interpretations leading to understanding cause and effect to be adequately informed on the issues of civic concern. With the educational establishment not fully able to deliver these skills, society must look at alternate sources and services for citizens to come to analytical and interpretive conclusions. Our ongoing studies point to the nexus of a solution being library sources and services. We have indications that adding value to library geospatial sources and services will improve the overall resolution of users' geospatial information needs in a complex society.

A short list of some additional work that should be conducted to further understanding of texts, maps, and user tasks with geospatial data and information includes:

- further exploring the role of different formats in adding value to geospatial information services and products as well as their interaction with tasks;
- developing a more comprehensive list of user geospatial information situations and tasks based on the actions and needs in real situations;
- exploring the expected and actual helps patrons receive from geospatial information;
- reviewing how libraries are currently meeting geospatial information needs, seeking out those that add value to basic data;
- assessing the use, usability, and usefulness of electronic geospatial data and the tools for displaying and manipulating such data in library contexts;

- investigating the transformations patrons make to effectively use current library products and services, and probing patrons' ideal solutions to geospatial information needs; and
- determining what role local, state, and federal governments as well as local organizations may play in supporting customization of geospatial data for local needs and making them accessible through public libraries.

REFERENCES

- Appleyard, D.; Lynch, K.; & Myer, J. (1964). *The view from the road*. Cambridge, MA: MIT Press.
- Armstrong, M.; Densham, P. J.; Lolonis, P.; & Rushton, G. (1992). Cartographic displays to support locational decision making. *Cartography and Geographic Information Systems*, 19(3), 154-164.
- Benbasat, I., & Dexter, A. S. (1986). An investigation of the effectiveness of color and graphical information representation under varying time constraints. *MIS Quarterly*, 10(1), 59-83.
- Bertin, J. (1981). *Graphics and graphical information processing* (W. J. Berg & P. Scott, Trans.). Berlin: Walter de Gruyter.
- Bieger, G. R., & Glock, M. D. (1986). Comprehending spatial and contextual information in picture-text instructions. *Journal of Experimental Education*, 54(4), 181-188.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: David Kay.
- Dervin, B. (1983). *An overview of sense-making research: Concepts, methods, results to date*. Unpublished paper presented at the International Communications Association, Dallas, TX.
- Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In J. D. Glazier & R. R. Powell (Eds.), *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited.
- Dervin, B., & Nilan, M. (1986). Information needs and uses. In M. E. Williams (Ed.), *Annual review of information science and technology*, 21, 3-33.
- Gluck, M. D. (1993). *Understanding performance in information systems: An investigation of system and user views of geographic information*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Gluck, M. D. (1995). Understanding performance in information systems: Blending relevance and competence. *Journal of the American Society for Information Science*, 46(6), 446-460.
- Gould, P., & White, R. (1986). *Mental maps* (2d ed.). Winchester, MA: Allen & Unwin.
- Guptill, S. C. (Ed.). (1988). *A process for evaluating geographic information systems* (U. S. Geological Survey Open-File Report 88-105). Washington, DC: USGPO.
- Larkin, J. H., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65-100.
- Lehrberger, J., & Bourbeau, L. (1988). *Machine translation: Linguistic characteristics of MT Systems and general methodology of evaluation*. Amsterdam: J Benjamin's Publishing.
- Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of the research. *Educational Communication and Technology*, 30(4), 195-232.
- Levin, J. R., & Lesgold, A. M. (1978). On pictures in prose. *Educational Communication and Technology*, 26(3), 233-243.
- Lindsay, P. H., & Norman, D. A. (1977). *Human information processing: An introduction to psychology*. New York: Academic Press.
- Lloyd, R., & Steinke, T. (1986). The identification of regional boundaries on cognitive maps. *Professional Geographer*, 38(2), 149-159.

- Lynch, K. (1960). *The image of the city*. Cambridge, MA: MIT Press.
- Monmonier, M. S. (1991). *How to lie with maps*. Chicago, IL: University of Chicago Press.
- Muehrcke, P. C. (1986). *Map use: Reading, analysis, and interpretation* (2d ed.). Madison, WI: JP Publications.
- Newby, G. B.; Nilan, M. S.; & Duvall, L. M. (1991). Toward a reassessment of individual differences for information systems: The power of user-based situational predictors. In J. M. Griffiths (Ed.), *Proceedings of the 54th ASIS annual meeting* (pp. 73-81). Medford, NJ: Learned Information.
- Nilan, M. S. (1985). *Structural constraints and situational information seeking: A test of two predictors in a sense-making context*. Unpublished doctoral dissertation, University of Washington, Seattle, WA.
- Nilan, M. S.; Newby, G. B.; Paik, W.; & Lopatin, K. (1989). User-oriented interfaces for computer systems: A user-defined online help system for desktop publishing. In J. Katzer & G. B. Newby (Eds.), *Proceedings of the 52nd ASIS annual meeting* (pp. 166-171). Medford, NJ: Learned Information.
- Schamber, L. (1991). *Users' criteria for evaluation in multimedia information seeking and use situations*. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Streeter, L. A.; Vitello, D.; & Wonsiewicz, S. A. (1985). How to tell people where to go: Comparing navigational aids. *International Journal of Man-Machine Studies*, 22(5), 549-562.
- Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14(4), 560-589.
- Thorndyke, P. W., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology*, 12(1), 137-175.
- Vanetti, E. J., & Allen, G. L. (1988). Communicating environmental knowledge: The spatial impact of verbal and spatial abilities on the production of route directions. *Environment & Behaviour*, 20(6), 667-682.
- Vaziri, M., & Lan, T. N. (1983). Perceived factors affecting driver route decisions. *Journal of Transportation Engineering*, 109(2), 297-311.

ROBERT LEE CHARTRAND

Emergency Preparedness and Response Challenges for Special Libraries

The challenges and opportunities facing special libraries of all types in our "republic of technology" are diverse and daunting. The evolving emergency preparedness and response missions, programs, and projects needed include the creation of a capability for providing geographic-oriented information in a number of critical application areas. This timely and trenchant topic deserves immediate attention by the cognizant leadership of those charged with library and information services.

At a time in our civilization when the watchwords include "cyberspace," "virtual reality," and "information society"—though these may be imperfectly understood—there still is a very real need to reexamine the role of traditional libraries, both public and private. The expanding role of information centers is a matter of current focus and concern, prompting growing discussion among the leadership in business and commerce, academe, and research institutions.

Among the key missions of private libraries, increasingly, are those related to emergency preparedness and response. It is gratifying that, during the past decade, there have been initiatives such as the preparation in 1987 of a topical issue in the journal *Special Libraries* (Chartrand, 1987) concentrating on a number of the crucial problems posed by emergency situations. The contributions made by librarians, information specialists, public officials, and emergency managers in that volume set the stage for some of the deliberations being undertaken at this clinic.

During a subsequent span of time, the rise in awareness regarding Geographic Information Systems (GIS) has been spectacular both within the private sector and in government at all levels. Indicative of this was a report entitled *Toward a Coordinated Spatial Data Infrastructure for the Nation* prepared by the National Research Council Mapping Science Committee (1993). Through a series of intensive interactions with eighteen federal departments and agencies over a two-year period, this group of senior specialists foresaw that "the scope of spatial data can be enormous, and spatial data can be important components of a wide variety of scientific, technical, and social disciplines and applications" (p. 1).

All levels of society are affected by the decisions made which involve dependency upon spatial data, including such applications as emergency

management (EM), real estate transactions, facility management, taxation, land-use planning, transportation, environmental assessment and monitoring, and research. To be prepared for these cascading demands on existing special libraries, their leadership must review those resources which now exist to meet these needs and determine promptly the types of responsive information—narrative, statistical, and graphic—that must be created or assembled.

If the elements of a decision support system using GIS (see National Research Council, 1990, p. 20) are examined, there are certain basic requirements to be met by special libraries for acquiring, encoding, storing, and maintaining geographic data. With the advent of new electronic capabilities for allowing library patrons—in corporations, “think tanks,” and universities, as well as the public at large—access to spatial and cartographic information have come new and exciting products and services. Needless to say, there is now a far better understanding on the part of the public of its right to know about such holdings (e.g., certain types of satellite photography, urban development maps). This has provided an added impetus to establishing and maintaining such graphic resources.

Just as the current administration is engaged in “reinventing government” with dozens of recommendations which include the creation of a National Spatial Data Infrastructure (NSDI), individual agencies are being charged with rethinking their emergency management roles. Within the former broader context, ten focal areas of concentration were identified for further analysis and selective enactment (Chartrand, 1994a):

1. Placement of the official emergency management “power center” so as to ensure full White House support for all relevant planning and response functions.
2. Periodic calculated revision of guidelines for departmental mission oversight (e.g., FEMA, DOD, EPA, etc.) by congressional entities.
3. Assertive monitoring of the emergency management community (twenty-seven federal agencies) in order to achieve fulfillment of “federal response plan” guidelines (FEMA, 1992).
4. Reevaluating the role of advisory groups in enhancing national EM effectiveness (e.g., FEMA Advisory Board).
5. Developing interactive guidelines and processes to strengthen substantially federal-state-local government action in all four phases of emergency management: prevention, mitigation, response, and recovery.
6. Improving the ways and means by which responsible governmental groups interact with cognizant private sector institutions (National Emergency Management Association, National Coordinating Council on Emergency Management, Alliance of Information and Referral Systems, Red Cross, libraries, “hotlines,” etc.).
7. Constructively enhancing planning procedures designed to facilitate interjurisdictional cooperation throughout all phases of emergency management.

8. Determining how best to develop and mesh into existing operations advanced information tools and techniques (computers, plotters, measurement sensors, telecommunications, multimedia display systems).
9. Effecting better working relationships among operational emergency management entities, libraries, other information resource centers, and the media in order to further public awareness, support, and conduct during crises.
10. Assessing recommendations contained in recent studies and reviews which set forth methodologies for heightening EM readiness, including the increased use of simulations and exercises (of varying degrees of sophistication) and more orientation and training for managers and "line" operatives. (p. 1)

For those responsible for improving the understanding and use of Geographic Information Systems as they are involved in disaster situations—whether in conventional research and reference settings, decision-making environments, or service to the public—it is important that the foregoing areas be recognized as to their substance and priority.

It is recognized that there are many who only wish that "disasters would go away," but the catastrophic events of recent years ranging from the Mount St. Helens eruption and the Loma Prieta earthquake to the central U.S. floods and Hurricane Andrew, to name only a few, have placed a very real burden on those who must strive to protect the lives and property of our citizens. More than a decade ago, then Representative Albert Gore, Jr. urged his fellow citizens to face the recurring depredations wrought by nature and sometimes our fellow man:

The subject of disasters is not one that many of us care to dwell on. Earthquakes, fires, assassinations, terrorist attacks, and nuclear melt downs are the stuff of Hollywood, and we like to keep it that way. As a result of this out of sight, out of mind ethic, our society is often ill-equipped to deal with emergencies when they do arise. Out of neglect, we fail to bring to bear either the technology or the management capability needed to respond quickly and effectively to the demands of stress situations. (U.S. Congress, 1984, p. 123)

In looking at some of the application areas where GIS capabilities need to be in place or "on tap" through linkages to special libraries, the following reflect actual "targets of opportunity":

- Use in the preparation of emergency scenarios, such as that featured in the CUSEC (Central United States Earthquake Consortium) report, *An Assessment of Damage and Casualties for Six Cities in the Central United States Resulting from Earthquakes in the New Madrid Seismic Zone* (FEMA, 1985).

- Availability of maps, charts, aerial and ground photography covering geographic and planimetric information for use by developers, regulatory agencies, citizens involved in real estate matters, and agencies (such as schools and churches) responsible for providing services that require specific information on population locations.
- Selective storage of GIS-type information on local and regional areas required by emergency services including police, fire, and medical forces that not only must take preventive action—such as moving impaired persons to safe sites—in the case of impending disasters, but also be positioned to cope with the aftermath of such events where familiar landmarks may be unrecognizable.
- Related to the previous support function is the provision of graphic information for public education purposes varying from lectures and seminars or documentary-type programs utilizing multimedia involving a geographic orientation, to the preparation of visual aids and hand-outs (hard copy or disk) for adult and student recipients.

In essence, these “GIS materials” have become a crucial part of the corpus of information required for the effective functioning of our public and private agencies and the citizens they serve. In addition, the larger corporations and small business communities, which comprise a significant segment of any town, city, county, or circumscribed region, need to have at their disposal such geo-oriented information. A seemingly localized disaster, such as those which occurred in production facilities at Bhopal or Chernobyl, may have the potential, as those in surrounding areas have found to their dismay, for a far broader and more devastating impact.

Another dimension to be considered is the desirability of standardizing such data in order to allow its sharing by those using different online computerized databases. One result of the 1988 Gatlinburg Symposium on Information Technology and Emergency Management was the identification of major deficiencies in the existing emergency management action arena. Noted among these in the ensuing volume, *Strategies and Systems for Disaster Survival* (Chartrand & Chartrand, 1988), were the lack of a taxonomy of core EM terms, the lack of standardized data definitions and computer interfaces, and the need for improved hardware and software compatibility. Another finding, of particular significance to those in the special library realm, dealt with the lack of useful and accurate reference works.

Thoughtful practitioner-scholars such as Lawrence Mondschein (1994), drawing upon the experiences derived from actual corporate investment in advanced planning and response systems, are having an effect on others who must face such challenges. In his perceptive paper entitled “The Role of Spatial Information Systems in Environmental Emergency Man-

agement," he describes in a case study the process by which Johnson and Johnson (J&J) selected an outside spatial information system to ensure achievement of its corporate goals: "(1) perform hazard assessments of chemicals stored on-site; (2) cross-train facility personnel with local authorities on emergency planning and response procedures; and (3) communicate accurate chemical inventory, location, and emergency health data during an actual incident" (p. 683).

Through its use of EISC—a PC- and LAN-based system widely used throughout this country and abroad—J&J has been able to draw upon this multifaceted capability which establishes and manages site-specific information on personnel; emergency resources; locations of schools, hospitals, and nursing homes; and environmental incidents or events. Chemical modeling is also a feature of this system created by Emergency Information Systems International with a database for storing worst-case scenarios that can be developed using chemical plume dispersion models (Chartrand, 1994c).

The acquisition of appropriate existing maps or their tailored counterparts is a problem not to be ignored, for in many instances, up-to-date community or facility maps are lacking. There may well be a need for high-resolution maps—ranging from 640×480 to 1024×768 pixels—with the choice being determined by the hardware available. U.S. Census Bureau TIGER line files may be useful in many cases. The transfer and sharing of both graphic and textual information among more than a dozen far-flung J&J locations, employing telecommunications software developed specifically for use with EISC, allows rapid review by such dispersed key staff. They can perform their requisite duties using packet radio, dial-up systems, cellular phones, or satellite transmission.

Many organizations are developing some form of "Emergency Operation Center (EOC)" which, when combined with a "special library," can serve to focus on and upgrade the corporate, academic, or other institutional information handling capability. As indicated earlier, there is a far better perception today among top officials, line managers, and rank-and-file operatives regarding the number of situations requiring information of a geographic nature. In addition to their vital role in preparatory actions for an actual emergency, these EOCs often may be the vortex for desk-top, or actual field-type, exercises to test the responsiveness of the organizational team. Experience has proven that the ability to incorporate realistic data into any type of "exercise" is contingent upon the imaginative exertions of those who plan and execute such training. Taking this kind of investment in time, materials, and staff one step further, the development of computer-manipulated models has been judged useful by diverse participants in congressionally-sponsored forums, hearings, and workshops during the past fifteen years.

In looking objectively at the capacity of many, if not most, special libraries to provide, upon demand, GIS-type information, it has to be conceded that there often is room for improvement on an array of fronts.

First, simply the collection of certain types of pertinent geographic information (maps, plats, charts, annotated photography, diagrams, building data, etc.). Second, the need for indexes which offer keyword, context term, or special geographic-point access to these holdings. Illustrative of this latter tool is the "overlapping polygon" developed by the IBM Corporation in the early 1960s (see Figure 3) (U.S. Congress, House of Representatives..., 1984, p. 231). Third, in order for the indexes to be easily used, the creation of a thesaurus of terms needs to be undertaken with every effort made to make such a tool helpful to all branches of the emergency response community. Finally, there is often a requirement for a "taxonomoy" or conceptual framework with standardized terminology and definitions for the focal field (which may well transcend "emergency management" per se). An example of this is *A Taxonomy of Human Services* developed by Georgia Sales (1991) at INFO LINE in Los Angeles.

Mention has been made of the importance of software selection, and a corollary to that process must be the hardware configuration which is to be employed. In today's fast-moving technological milieu, it often seems that no sooner is one "system" put in place and learned, with the traumas and time-consuming efforts which accompany that integration into the existing work environment, than another "better, faster, cheaper, more 'user-friendly'" one is being marketed. The best way to minimize or remove—at least for the time being—this threat to managerial effectiveness and personnel morale (and sanity), is through the inception of a multiple-year strategic plan with modular annual increments of meaningful detail for the near future and acceptable lesser specificity for the later time phases.

Cognizance of the importance of enhanced emergency preparedness and response is reflected on the contemporary scene in a series of high priority initiatives assertively supported by Vice President Gore (Chartrand, 1994b).

1. Redirection of EM efforts to encompass an "all-hazards" approach to ensure protection of life and property.
2. Reorganization of FEMA under the directorship of James Lee Witt, a highly qualified professional in the field.
3. Determination to significantly enhance federal support to state and local organizations, and explore federal-state technology partnerships where applicable.
4. Development of a stronger understanding and utilization of the critical role of high technology in disaster prevention, mitigation, response, and recovery.

In conclusion, it is the responsibility of all governmental and societal components whose missions include emergency preparedness and response to rise to this challenge. Here, then, is an unprecedented opportunity for the library and information science community—known for its innovative approaches to meeting and overcoming the stressful demands endemic in this turbulent Age of Information—to play an essential role in evolving a cohesive plan for marshaling all requisite information resources that will ensure an optimum preparedness and response posture for the nation.

REFERENCES

- Chartrand, R. L. (Ed.). (1987). Emergency management [Special issue]. *Special Libraries*, 78(2).
- Chartrand, R. L. (1994a). *Emergency management update: Initiatives and opportunities*. Unpublished paper presented at the 5th National Institute for Urban Search and Rescue (NI/USR), Town Meeting and Information Systems Conference, March 14-15, 1994, Rockville, MD.
- Chartrand, R. L. (1994b). *Round Table on emergency management* (The Naples Institute, December 6-7, 1994). Naples, FL: Naples Institute.
- Chartrand, R. L. (Ed.). (1994c). *Your all-hazards solution*. Rockville, MD: EIS International.
- Chartrand, R. L., & Chartrand, K. C. (Eds.). (1988). *Strategies and systems for disaster survival* (The Symposium for Information Technology and Emergency Management, May 4-6, 1988, Gatlinburg, TN). Rockville, MD: Research Alternatives.
- Federal Emergency Management Agency. (1985). *An assessment of damage and casualties for six cities in the central United States from earthquakes in the New Madrid seismic zone*. Washington, DC: FEMA.
- Federal Emergency Management Agency. (1992). *The Federal response plan for Public Law 93-288, as amended*. Washington, DC: USGPO.
- Mondschein, L. G. (1994). The role of spatial information systems in environmental emergency management. *Journal of the American Society for Information Science*, 45(9), 678-685.
- National Research Council. Mapping Science Committee. (1990). *Spatial data needs: The future of the National Mapping Program*. Washington, DC: National Academy Press.
- National Research Council. Mapping Science Committee. (1993). *Toward a coordinated spatial data infrastructure for the nation*. Washington, DC: National Academy Press.
- Sales, G. (Ed.). (1991). *A taxonomy of human services* (2d ed.). El Monte, CA: Information and Referral Federation of Los Angeles County, Inc. and Joliet, IL, Alliance of Information and Referral Systems.
- U. S. Congress. House of Representatives. Committee on Science and Technology. Subcommittee on Investigations and Oversight. (1984). *Information technology for emergency management*. Washington, DC: USGPO.

Using Geographic Information Systems for Estimating and Profiling Geographic Library Market Areas

A definition of legal service or market area is difficult for public library management due to limited available data regarding user residence, and because people may cross service lines for any number of reasons. Yet an accurate estimate and a subsequent socioeconomic profile of the geographic market to be served (market analysis) is essential in order to provide unique community-based services and materials. Geographic information system (GIS) software can facilitate library market analysis by graphically estimating geographic boundaries and analyzing socioeconomic characteristics within prescribed markets in one online environment. This discussion illustrates the utility of GIS in estimating and profiling library markets. The Evansville-Vanderburgh County public library system is used to provide realistic library market analysis situations.

INTRODUCTION

A definition of legal service or market area¹ is difficult and complex for public library management to establish for two major reasons: (1) library managers may not know where the majority of users reside due to limited knowledge and data collection, and (2) because people may cross service lines for any number of reasons (Van House et al., 1987). Historically, the library profession develops market areas from the "inside-out" and the "outside-in." The "inside-out" approach would be quantitative standards mandating a one mile service area (Eastman, 1911) and a half-mile service area for children (ALA, 1956), or optimal service areas for one or one and one-half miles in urban areas (ALA, 1943).

These parameters are based upon assumptions that not only do the majority of users or potential users live within these radii, but the radii indicate the maximum distance a user will travel to the facility. Also, when branches were placed closer together before the advent of the automobile, the one mile rule was established almost by default. Distance between library facilities and its effect on use, and subsequently upon market size, has been the subject of study (Grundt, 1968; Schlipf, 1973; Coughlin et al., 1972; Getz, 1978; Hayes & Palmer, 1983).

Size or average radius in miles of the market area for library facilities has also been the focus of study (Shaughnessy, 1970; Coughlin et al., 1972;

Schlipf, 1973; Getz, 1980; Freestone & Beard, 1981; Koontz, 1990; Shoham et al., 1990). A summary of studies of market area size nationwide indicated the average metropolitan library market has a radius of two miles (Palmer, 1981).

All the quantitative standards were officially dropped in the 1960s (Public Library Association, 1967) and the "outside-in" approach began. A summary of currently used methods includes: justifying metropolitan branch markets with a potential population of 30,000 served and the nearest branch is three to four miles away (Wheeler & Goldhor, 1962); assigning each library branch certain census tracts; using existing local government planning zones; or conducting a sample of registration or circulation records, plotting the results on a large map, and drawing largest percentages of populations proximate to each branch location (Van House et al., 1987). The latter sample survey method reinforces the estimate of the market area thereby providing a more precise definition.

It is agreed that all approaches, at best, are rough estimates, and that some communities' markets are easier to define than others. For example, a rural area with one library serving a small county would simply have county boundary lines, while an urban area—e.g., the Los Angeles County Public Library with sixty or more branches—would be more complex, involving many more considerations such as distance to nearest facility, transportation networks, topological boundaries (airports, parks, etc.), and the socioeconomics of the potential user market.

Across America there are an estimated 16,000 public library markets. By and large, these are branch markets. Estimating and profiling the geographic market area is the first step in analyzing the market of people within the area to be served. This market profile of the community served is, of course, what all services and material offerings are supposed to be based upon (Public Library Association, 1979). Yet the library manager, through lack of data and tools, is often forced to haphazardly estimate the market area. Accurate measurement and subsequent definition of the market area then, must become a priority for library managers and researchers alike to ensure that use of the library is optimized in communities throughout America.

NEW TECHNOLOGIES FACILITATE LIBRARY MARKET ANALYSIS

A new technology can assist in the market analysis process. Geographic information systems have burgeoned within the past ten years. A GIS, as a computer-assisted system, is used for the capture, storage, retrieval, analysis, and display of spatial data. Spatial data describe location

and geographic relationships among things and events. For the public library, this would describe the relationship between library use and the geographic market of users served.

Graphical maps are utilized in many GIS applications. It is estimated that over 80 percent of governmental decisions involve spatial data, making quick access to spatial data crucial to effective government operations. GIS technology is an attractive and efficient alternative to the manual processes of spatial analyses traditionally performed by public and private sector managers. The utility of GIS to estimate and define library market areas will be discussed and illustrated in this article.

PROBLEMS WITH, AND SOLUTIONS FOR, USING GIS

Problems

At present, there are some problems with GISs that should be noted before application is illustrated.

- *Data accuracy and error*—in a spatial setting, data error is especially hazardous. Setting map layers on top of each other, which are either gathered at different projections or collected at a rate of error from 5 to 100 feet, is serious business. Data can also be mislabeled. The amateur user of a GIS must be guided by any available expertise.
- *Data availability and procurement*—Desired data may not always be available. Fortunately, because of the growing availability of computers in libraries, data collection regarding library use can be automatic. The *Public Library Data Service Statistical Report* (PLA, 1988-1994) is a summary of important field data. The U.S. Census Bureau market data have been online since 1980. This type of summary data collected at higher levels and distributed widely saves tremendous cost to users. Data procurement is the most expensive aspect of GIS.
- *Very steep learning curve*—GIS is still difficult to use. Unfortunately, it is something you need to work with almost daily to be familiar with all aspects of the technique. But as counties and other government entities acquire GIS, the library manager may simply need to understand the analytical capabilities of GIS and know what questions to ask.
- *Expensive hardware and software*—until recently, GIS could not be purchased for under several thousand dollars. Now software packages are being produced that cost under \$1,000.

Solutions

Problems associated with GISs are being solved, and access by library managers is more realistic as more users come to the GIS market thereby creating a need for more user-friendly GISs.

New Computerized Data Products. New computerized products from the U.S. Census Bureau that are utilized within a GIS framework are inexpensive and widely available. TIGER (Topologically Integrated Geographic Encoding and Referencing) line files are computerized computer-readable maps containing graphical linework images of nearly every street in the United States. TIGER files are inexpensive and ultimately usable by most GIS software. In addition, all the data gathered from the decennial census can be referenced to lineworks depicting geographic features, such as census tracts,² in the TIGER file. Thus, these socioeconomic data can be integrated into a GIS database using TIGER line files as a basis. A market profile that includes relevant census data elements such as age, race, sex, income, education levels, etc., for a given geographic area, can be produced from the GIS database. Library use data can be entered as another layer and viewed graphically within an analytic framework with other data (e.g., census, market boundaries, etc.).

While geographic information systems are traditionally used in areas of land management, natural resources, and highway planning because of the inherent spatial nature of these entities, private and public sector managers are beginning to see the utility of GIS for demographic market analysis. Many counties are now purchasing, or have purchased, GIS software for land management and planning. Agencies, such as libraries, desiring demographic analysis, can benefit from such local GIS purchases by developing their own applications for use with the local GIS software and hardware.

MARKET ANALYSIS

As discussed earlier, before any decision can be made regarding size and location of facility and materials and services that are offered, an estimate and profile of the library market area must be derived.

This first aspect of the market analysis consists of three questions:

1. How large is the current market (current geographic market and population size estimate)?
2. Who is the market (market definition and profile)?
3. What is the likely future size of the market (market forecasting)?

There are two remaining tasks within the market analysis that must occur after measurement and identification of the market area. The second is market segmentation, which is the process of determining the main groups to be served within the market area. The third is consumer analysis, determining the characteristics of users, specifically their needs, percep-

tions, preferences, and behavior (Kotler, 1982). The second and third steps are outside the scope of this discussion with the exception of identifying the demographic characteristics of users and the levels of library use. A GIS can greatly facilitate all tasks of the market analysis through the four powerful capabilities discussed later.

GIS ANALYTICAL CAPABILITIES FOR MARKET ANALYSIS

A useful classification of GIS analytical capabilities is provided by Thrall & Elshaw (1993). They categorize these capabilities into description, explanation, prediction, and judgment activities.

- *Description* documents and describes the spatial landscape (e.g., where are the census tracts with more than 35 percent of the population between the ages of 0 and 18, or which census tracts have a population with more than a 50 percent black population?).
- *Explanation* analyzes the phenomena that are found in the description phase (e.g., library usage is low for a particular branch because 20 percent of the population is on public assistance, 37 percent are 65 or older, and only 10 percent of the households are occupied by couples with children). Research indicates that all these characteristics are usually associated with low use markets (Koontz, 1990).
- *Prediction* uses modeling and statistical analysis to predict changes in a particular variable based on changes in other variables (e.g., systemwide library use will increase 20 percent when a new branch opens in a quadrant which is at present unserved).
- *Judgment* (or prescription) uses the findings of the first three types of analysis to prescribe an action (e.g., a long-range systemwide facility location and service plan).

For purposes of this presentation, the analytical capabilities of description and explanation will be used to estimate, measure, and define and profile library market areas.

GIS FOR MARKET ESTIMATION AND PROFILING

Representative uses of a GIS in a public library arena setting for market estimation and profiling by a library manager will now be presented. The public library setting that is used is the Evansville-Vanderburgh County, Indiana, public library system. The director provided the author with data from the county's library system.

A REVIEW OF THE EVANSVILLE-VANDERBURGH COUNTY PUBLIC LIBRARY SYSTEM

The Evansville-Vanderburgh County Public Library System has a central library, seven branches, young adult literacy outreach, homebound delivery, nursing home collections, and talking books services. An estimated population of 167,000 is served.

The East and West branches were the first libraries built in the newly formed Evansville Public Library in 1913. The Stringtown branch opened in 1939, following suburban growth outward from the downtown. Like many library systems across the country, the 1960s and 1970s were a time of expansion and construction due to the availability of funds provided by the Library Services Act in 1956 and the Library Services Construction Act in 1964. On the third level of growth outward from the inner city ring and beyond Stringtown, the McCollough Library (1965), the North Park Library (1968), and the Oaklyn Library (mid-1970s) were built in response to sprawling urban growth and the availability of funds.

The Red Bank Library was opened in 1991 in the University Shopping Center and draws users from farther distances due to traffic from diverse groups coming to university functions. A new branch for the northeast quadrant is also being considered and will be discussed in a later section on market definition.

MARKET ESTIMATE AND MEASUREMENT

As discussed earlier, geographic determination of market area is difficult. This is true not only for library managers but also for any manager of a store or service that is traveled to by its customers or users. Retailers have often used radii for store markets in order to generalize and identify key characteristics within the circles. Yet marketing consultants readily agree that polygons or specified market shapes are, and should be, unique to each location. These irregular service or market boundaries are more difficult to determine (Reid, 1995). A GIS facilitates a variety of ways that market boundaries can be defined, including irregular noncircular market areas. These methods, when employed by the library manager, basically reflect the more modern "outside-in" approach. Five graphical approaches to market measurement are discussed below. Market definition in conjunction with these measurements will be discussed and illustrated next. Definition describes who (socioeconomic census data) lives within the designated geographic areas.

Approaches to Market Measurement

There are five approaches to market measurement:

1. *Assigning each library branch a certain number of census tracts.* The Evansville-Vanderburgh Public Library assigns census tracts to each branch.
2. *Determining block groups³ within the census tracts assigned to each branch market.* This provides further opportunity for analysis of smaller portions of the geographic library market area.
3. *Determining a branch market through overlay of zip code boundaries.* The value of zip code boundaries is important when a library records user zip codes, identifying where users reside.
4. *Determining branch markets by assigning equal portions of the population to the nearest existing facility.* This is a modeling⁴ technique, location allocation which simply assigns each member of the population to the nearest facility, in this case a library branch.
5. *Determine a branch market by assigning a certain mile radii to be served.* This is a generalized approach, usually used by managers for a standardized point of comparison of key features. This is the average market radius of a metropolitan branch market (Palmer, 1981).

While these market measurement methods are not exhaustive, they show the dynamics and versatility of GIS in measuring and determining geographic library markets.

Market Definition and Profile

In order to answer who resides within these geographic areas, such as census tracts, block groups, zip codes, equal areas and circles, and ultimately, within the library market area, however defined, U.S. Census data must be attached and geo-coded to these areas. These data form another layer that can be displayed in the GIS environment.

There are literally hundreds of variables to select from in the census data. Each manager must know which variables are relevant to their user or customer. Research in the library field (Koontz, 1990) recognizes certain broad variable groups that are strongly associated with library use. These variables are also recorded by the U.S. Bureau of the Census. The nine broad variable categories include: (1) population (Palmer, 1981); (2) sex (Zweizig, 1973); (3) race (International Research Associates, 1963; Koontz, 1990); (4) age (Kronus, 1973; Hayes & Palmer, 1983); (5) family life cycle (represented by the census data categories of households with social security) (Hayes & Palmer, 1983), public assistance (Marketing Institute, 1988), or a female head of household⁵); (6) owner occupied housing (Zweizig, 1973); (7) income (Coughlin et al., 1972; Schlipf, 1973; Bennett & Smith, 1975; Getz, 1978; Van House, 1983); (8) education (Kronus, 1973; Zweizig, 1973; Hayes & Palmer, 1983); and (9) vehicles per housing unit (D'Elia, 1980; Gallup, 1976).

Thirty-three variables among the broad categories are initially measured and analyzed for each facility's market area (see Appendix A for a full listing of these variables).

The five types of geographic market areas, as measured by various boundaries, include census tract, block group, zip code, equal area and radius, and will define the market population of the Red Bank library branch market area with the use of GIS. Because census tracts and the block groups within represent the same geographic area in total, the numbers associated with each variable will add up equally. The benefit of block group information is to scrutinize and analyze a smaller geographic area.

Zip code information, in general, provides an "inside-out" look at who uses the library. For example, if user registration includes zip code information, a graphical display could be developed which would be useful in determining where the largest or smallest percentage of users live, how far most travel to the library, or perhaps why some travel to another library rather than the closest (e.g., special holdings, location, hours, etc.). Zip code data can also be used for direct mail purposes to publicize library services and programs to target populations—e.g., announcing an English literacy program to a heavily Hispanic neighborhood. This would be a benefit of using census block group, zip code, and library data in combination and graphically displayed in one environment.

Equal market area modeling (assigning population in equal amounts to nearest branch) offers an opportunity to see projected gaps or overlaps in service.

The radius, when applied at differing increments, can be used more successfully when it is known where percentages of users live. For example, do over 75 percent live within a mile or where does the fall-off occur? The circle also offers a standard point of comparison of key features within markets.

Each of the above methods is complementary in strength in providing a general description of the market area. These methods should be considered the "first cut" at market estimate and subsequent profile. For a more precise estimate and profile, a survey of user residences should be made. Each facility in the system has an impact on all others, and a review of all facility markets should be made. A review of gaps and overlaps in service is then possible.

Further Market Definition

The utility of GIS in market definition, and the enormous amount of information that can be provided for any geographic area, is illustrated in the discussion above. But more precise information can also be obtained. The following scenarios will illustrate what information library management may need to provide.

Scenario One. In preparation for planning the summer reading program, the library systems' children's librarian wants to identify the percentages of census tracts with over 35 percent of children 0-18 in age.

Scenario Two. A grant opportunity is available from a private foundation fund to provide literacy outreach services in majority black neighborhoods. The grant writer for the library would like to identify those census tracts within any library market that are over 50 percent black. The Eastside library appears to serve this group.

Scenario Three. It is time for the annual budget request to the county commission. The library director simply wants to know which library markets meet the higher quartile of annual per capita circulation as reported in the *Public Library Data Service Statistical Report* (Public Library Association, 1992). The higher quartile reported for systems serving 100,000 to 249,000 is 8.3 per capita. The Evansville-Vanderburgh County Public Library serves 167,000.

Scenario Four. The library system director is considering establishing a new branch in the northeast quadrant that is presently unserved. Industrial growth is presently taking place, and residential growth is predicted within the next five years. The director needs to know what is the current population, and what is predicted (this will be obtained from the local planning department), and also needs a review of the topography and major roads in order to review the proposed site at the junction of Interstate 64 and Morgan Avenue.

The library's criteria for new branches includes a population base of at least 25,000 to 30,000 within a two mile radius. Note that, at present, the population is only 4,900 within the two mile radius. The planning department projects approximately 10,000 more people by year's end. This is 10,000 short of the prescribed number recommended for a branch.

These four scenarios illustrate the dynamics and power of using a GIS for market definition and library market profiling. To have a myriad number of data sets collapsed into one environment for all types of analysis places powerful and much needed tools in the hands of the library manager.

SUMMARY

While stone walls do not a prison make nor iron bars a cage, yet neither do circles or preset lines necessarily make a library market area. A GIS provides the library manager with an opportunity to better measure market areas and subsequently define those markets in a complex dynamic online environment. The goal is one of offering the best possible

materials and services to potential and actual public library users who are guided by community standards and the policies and procedures of the library.

As the twenty-first century looms, the public library is once again being challenged by new media and information access technology. While digital online libraries-without-walls are continually discussed, the geographic place of the library-with-walls within the community is simply heightened by these discussions. Access to burgeoning online data can only be possible if public libraries as unique public information agencies assure equitable access to the nontechnical elite and the information poor by being strategically placed geographically. A GIS will play a key role in helping public library managers meet this critical goal.

APPENDIX A

Census Market Variables

Persons - population

Families

Households

Females

Whites

Blacks

Indians

Asians

Hispanics

Age 3 to 4

Age 7 to 9

Age 10 to 17

Age 0 to 18

Age 18 to 19

Age 20 to 29

Age 30 to 44

Age 45 to 59

Age 60 to 64

Age 65 to 74

Age 75 and over

Education up to grade 9

Education grade 9 through 12th

Education - High School Graduate

Education some college

Education AA degree

Education Bachelor's degree

Households with earnings

Households with social security income

Household with public assistance

Number of houses

Number of owned houses

Number of owner occupied houses

Households with one vehicle

P = Percentage of the above groups

Temp1 field is library use per capita

NOTES

- ¹ A market area is the geographic area from which a library draws most of its users (Ghosh & McLafferty, 1987, pp. 12-13).
- ² A census tract is a large neighborhood generally with a population between 1,500 and 6,000 (Myers, 1992, p. 16).
- ³ A block group is a smaller subdivision within a census tract (Myers, 1992, p. 18) providing further geographic definition of a market area.
- ⁴ A model is a representation of relevant properties of reality, and models are usually mathematical. Variables are identified, defined, and measured. Models can trace various alternatives in the decision-making process. The relationships within models can describe and explain the past, provide control for the present, and enable prediction. The more relevant details there are in a model, the more opportunity there is for a true representation of reality (*Library Planning and Decision Making Systems*, 1974, pp. 76-85).
- ⁵ These variable categories are used to assess the importance of these lifestyle situations. Each may also serve as a measure of low income.

REFERENCES

- American Library Association. Commission on Post War Planning. (1943). *Post war standards for public libraries*. Chicago, IL: American Library Association.
- American Library Association. Coordinating Committee on Revision of Public Library Standards. (1956). *Public library standards: A guide to evaluation with minimum standards*. Chicago, IL: American Library Association.
- Bennett, W. D., & Smith, B. W. (1975). The correlates of library patronage distance decay. *East Lakes Geographer*, 10, 33-44.
- Coughlin, R. E.; Taieb, F.; & Stevens, B.H. (1972). *Urban analysis for branch library system planning*. Westport, CT: Greenwood Publishing Co.
- D'Elia, G. (1980). The development and testing of a conceptual model of public library user behavior. *Library Quarterly*, 50, 410-430.
- Eastman, L.A. (1911). Branch libraries and other distribution agencies. *Preprint of Manual of Library Economy*, chap. 15. Chicago, IL: American Library Association.
- Freestone, R., & Beard, D. (1981). A note on the library service area. *Australian Library Journal*, 30(November), 129-133.
- Gallup Organization, Inc. (1976). *The role of libraries in America*. Frankfurt, KY: Kentucky Department of Library and Archives.
- Getz, M. (1978). *The efficient level of public library services* (Working Paper No. 55). Cambridge, MA: The Joint Center for Urban Studies of M.I.T. and Harvard University.
- Getz, M. (1980). *Public libraries: An economic view*. Baltimore, MD: Johns Hopkins University Press.
- Ghosh, A., & McLafferty, S. L. (1987). *Location strategies for retail and service firms*. Lexington, MA: Lexington Books.
- Grundt, L. (1968). *Efficient patterns of adequate library service in a large city: A survey of Boston*. Urbana-Champaign, IL: University of Illinois.
- Hayes, R., & Palmer, E. S. (1983). The effects of distance upon use of libraries: Case studies based on a survey of users of the Los Angeles Public Library Central Library and branches. *Library Research*, 5, 67-100.
- International Research Associates. (1963). *Access to public libraries: A research report* (prepared for the Library Administration Division, American Library Association). Chicago, IL: American Library Association.
- Koontz, C. M. (1990). *Market-based modelling for public library facility location and use-forecasting*. Unpublished doctoral dissertation, Florida State University.
- Kotler, P. (1982). *Marketing for non-profit organizations* (2d ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kronus, C. L. (1973). Patterns of adult library use: A regression and path analysis. *Adult Education*, 23, 115-131.

- Library Planning and Decision Making Systems*. (1974). Cambridge, MA: MIT Press.
- Marketing Institute. (1988). *A better way: Assessing and planning library services to meet community needs, the Robert Homes Model* (prepared for the Chicago Public Library). Chicago, IL: Marketing Institute, Inc.
- Myers, D. (1992). *Analysis with local census data: Portraits of change*. London, England: Academic Press.
- Palmer, E. S. (1981). The effect of distance on public library use: A literature survey. *Library Research*, 3, 315-334.
- Public Library Association. (1967). *Minimum standards for public library systems 1966*. Chicago, IL: American Library Association.
- Public Library Association. (1979). Goals, guidelines and standards committee. *The public library mission statement and its imperatives for service*. Chicago, IL: American Library Association.
- Public Library Association. (1988-1994). *Public library data service statistical report* (prepared by the Public Library Association, a division of the American Library Association). Chicago, IL: American Library Association.
- Reid, H. (1995). Defining trade areas or trying to see the invisible. *Business Geographics*, 3, 46-47.
- Schlipf, F. A. (1973). *The geographical distribution of urban public library use and its relationship to the location of branch libraries*. Unpublished doctoral dissertation, University of Chicago.
- Shaughnessy, T. W. (1970). *The influence of distance and travel-time on central library use*. Unpublished doctoral dissertation, Rutgers University.
- Shoham, S.; Herskovitz, S.; & Metzger, D. (1990). Distribution of libraries in an urban space and its effect on their use: The case of Tel Aviv. *Library & Information Science Research*, 12, 167-181.
- Thrall, G. I., & Elshaw, S. (1993). Commercial data for the business GIS. *Geo Info Systems*, 3, 63-68.
- Van House, N. A. (1983). A time allocation theory of public library use. *Library & Information Science Research*, 5, 365-384.
- Van House, N. A.; Lynch, M. J.; McClure, C. R.; Zweizig, D. L.; & Rodger, E. J. (1987). *Output measures for public libraries* (prepared for the Public Library Development Project (2d ed.)). Chicago, IL: American Library Association.
- Wheeler, J. L., & Goldhor, H. (1962). *Practical administration of public libraries*. New York: Harper & Row.
- Zweizig, D. L. (1973). *Predicting the role of library use: An empirical study of the role of the public library in the life of the adult public*. Unpublished doctoral dissertation, Syracuse University.

Implementing GIS in the Public Library Arena

The rate of introduction of Geographic Information Systems (GIS) into public libraries is accelerating. This study investigates the experiences of all types of libraries that have either successfully or unsuccessfully introduced GIS into a library environment. From librarians' experiences, it is apparent that the most important consideration for successful GIS introduction into a library is adequate staffing and staff training. From this and other information collected during this study, a decision flow-chart was developed to help public librarians evaluate the type of GIS services that should be provided in any given library environment. The Public Library Users Geographic Information System (PLUS+GIS) research project, based in Florida, is a project focused on developing methodologies to assist public libraries in introducing GIS to their patrons.

INTRODUCTION

Geographic information systems are starting to appear in more and more libraries. GIS hardware, software, and the accompanying data are usually fairly expensive, however. The level of user support that may need to be provided by librarians for such systems may also be unacceptably high. With the increasing interest in linking GIS and libraries, a critical examination of the factors leading to successful versus unsuccessful implementation of GISs in a library environment would be helpful to other libraries contemplating such a decision.

This research paper will provide an overview of Geographic Information Systems in a variety of library settings and analyze the most critical factors leading to successful or unsuccessful implementation of a GIS in libraries. This information will be used to develop a model for implementing GIS in a public library environment.

GEOGRAPHIC INFORMATION SYSTEMS

Geographic information systems contain information stored in digital form that can be referenced to some point (e.g., storage well), line (e.g., river), or area (e.g., census tract), on Earth (or even in outer space). A formal definition for a GIS that has been adopted by the National Center for Geographic Information and Analysis is:

JUE/IMPLEMENTING GIS IN THE PUBLIC LIBRARY ARENA

A geographic information system is a computerized database management system for capture, storage, retrieval, analysis, and display of spatial (locationally defined) data.

The value of a GIS is the ability to store and relate attribute data (e.g., census data) with the geographic features (e.g., the census tracts) and view them relative to other geographic features.

GIS AND LIBRARIES

Libraries, whether they are research, public, state, or corporate, have become interested in GIS software and data because 80 percent of governmental information has a spatial aspect to it (Huxhold, 1991). More and more, these data are being distributed primarily or exclusively in electronic format. For instance, the TIGER (Topologically Integrated Geographic Encoding and Referencing) files developed by the U.S. Bureau of the Census and distributed on CD-ROMs can be processed into a digital map showing the appropriate census information for the entire United States.

Given the electronic nature of GIS, what is the role, if any, of a library in providing access to such computer software and its associated data? Several recent articles have addressed the issue of the proper role of a librarian in this new digital environment (Allen, 1993; Kollen & Baldwin, 1993; McGlamery & Lamont, 1994; Wong, 1993). Major issues that need to be considered by any librarian attempting to implement a GIS within their library environment include:

1. Is the library a "library" or a "laboratory" (Kollen & Baldwin, 1993)?
2. What is the role of the librarian? Is it merely to be an access facilitator or to also include the roles of collector, selector, and finder of spatial data (Allen, 1993) or perhaps even "map-maker?"

The answers to such questions have varied widely among the different libraries that have made GIS and spatial data available to their patrons. An examination of libraries that have actually placed GIS into a library setting will help provide additional insights into these questions.

GIS LIBRARY PROJECTS

The decreasing costs of computer hardware in conjunction with the increasing power of this same hardware have made it possible for libraries to consider providing GIS access for their clients. During the last few years, several significant projects involving Geographic Information Systems and libraries have been undertaken.

The most widespread and coordinated project is that by the Association of Research Libraries (ARL) and the Environmental Systems Research Institute (ESRI). The program currently includes about seventy research libraries. The goals of the project are:

1. to introduce GIS to a variety of libraries;
2. to develop GIS expertise in the library community;
3. to encourage connections and communication between GIS users and government agencies;
4. to promote education and research through improved public access to information;
5. to initiate library projects to explore new applications of spatially referenced information; and
6. to evaluate the introduction of these GIS services into research libraries (McGlamery & Lamont, 1994).

The ARL project was begun in the spring of 1992. The project has involved providing software, data sets, and training to librarians. An e-mail discussion list has also been established to facilitate communication among the ARL libraries participating in this project.

A more recent and ambitious project was begun in 1994 that involves libraries and spatial data. The Alexandria Project is one of the six digital library projects chosen to be funded by the National Science Foundation during 1994. The goal of this project is to develop a user-friendly digital library system that supports both textual and spatially indexed sources of information and that is scalable on a national level. One of the main initial outputs of this project will be a distributed testbed system appearing to the user as a single library. The research and development team includes not only computer scientists but librarians as well, including the Map and Image Laboratory of the University of California at Santa Barbara Library, the library of the State University of New York at Buffalo, the U.S. Library of Congress, the library of the U.S. Geological Survey, and the St. Louis Public Library. This project is based and administered at UC Santa Barbara.

Finally, a number of projects involving public libraries and Geographic Information Systems have also begun recently. Unlike research libraries, there is no coordination among these projects but rather a number of independent projects undertaken by each particular library system. Oftentimes this has been done in conjunction with an outside entity. Perhaps the best-known public library GIS project is the one at St. Louis Public Library and its Electronic Atlas project using ESRI's ArcView software (Kofron & Watts, 1993). Other public library systems with GISs at varying stages of implementation include Bellevue Regional Library in

Washington, Boston Public Library in Massachusetts, Cambridge Public Library in Massachusetts, Charlotte Public Library in North Carolina, Cleveland Public Library in Ohio, District of Columbia Public Library in Washington, D.C., Leon County Public Library in Florida, Monmouth County Library in New Jersey, New York Public Library in New York, Newport Beach Central Library in California, Ontario City Library in California, and the Seattle Public Library in Washington. By the time this article appears in print, the number will be significantly higher. As recently as just seven months ago (August 1994), McGraw (1994) could only locate four public library branches with interest in, and possession of, GIS programs.

RESEARCH METHODOLOGY

As part of the research in developing a model for implementing Geographic Information Systems in a public library environment, a list of libraries that were known to have attempted to introduce GIS into their own library was developed. These libraries included not only public libraries but also research libraries and state libraries as well. The names of these libraries were obtained by three primary means:

1. a list of the ARL GIS Literacy Project participants;
2. a review of popular GIS literature (e.g., GIS World, Government Technology) over the last five years; and
3. a query of all librarians contacted about other GISs in a library setting of which he or she may have knowledge.

This generated a total list of eighty-eight libraries. Sixty-six of the libraries (75 percent) were participants in the ARL project. During the months of February and March of 1995, phone calls were made to these eighty-eight libraries. A total of seventy-three individuals (83 percent) were successfully contacted. Not all libraries contacted had actually implemented a GIS in their library for one reason or another. Of the seventy-three libraries, sixty (82 percent) had a GIS.

REVIEW OF LIBRARY EXPERIENCES IN IMPLEMENTING GIS

Once contacted, participants were asked to discuss several aspects of their Geographic Information Systems environment and their experiences in implementing GIS (see Appendix A for a list of the discussion points). In some instances, not all questions were asked because the questions were clearly not applicable given earlier responses or this researcher had prior knowledge about some of the answers to the questions.

JUE/IMPLEMENTING GIS IN THE PUBLIC LIBRARY ARENA

The responses were entered into a database for further analysis. Analyses were performed on the collected responses in several ways:

1. all respondents as a whole (sixty-four libraries total);
2. respondents from the public library arena only (eleven libraries total);
3. respondents whose GIS implementation experiences could be classified as outstanding or exemplary (seventeen libraries total);
4. respondents whose GIS implementation experiences could be classified as disappointing (fifteen libraries total).

The category of outstanding/exemplary GIS implementation is based completely on this researcher's subjective evaluation of the respondent's success in introducing GIS to their library patrons. The respondents were not asked to evaluate their own efforts to implement GIS in their library. The library's primary clientele could have been college students or the general public.

The category of a disappointing Geographic Information System implementation is based on the respondent's voluntary evaluation of their own library's experience. Remember that the respondents were not asked to evaluate their own efforts to implement GIS in their library. As a result, this library category may better measure an individual respondent's own expectations than an actual difference relative to other libraries with "average" GIS implementation experiences.

It was hoped that this breakdown in respondent categories would help provide a clearer picture of factors likely to lead to successful versus unsuccessful GIS implementation as well as provide a clearer picture of any problems that may be unique to public libraries.

RESULTS OF PHONE SURVEY

A number of questions were asked of the respondents. The only ones that will be discussed are those that deal with a library's likelihood of successful GIS implementation. Table 1 divides the libraries into four groups. The greatest barriers to GIS implementation as identified by each respondent are given in the columns labeled from one to eight. An individual could identify more than one problem, so the totals across a row do not sum to 100 percent.

Regardless of the library grouping, the biggest problem is the large amount of training required for library staff to successfully support the GIS environment. This is due to the complexity of the GIS software. This is the biggest problem even for those libraries that were identified as having the greatest success in implementing GIS in their library. Even after staff is trained, the amount of time required for librarians to support GIS users is high and considered to be the second highest-ranking problem by librarians as a whole.

JUE/IMPLEMENTING GIS IN THE PUBLIC LIBRARY ARENA

TABLE 1
BARRIERS TO GIS IMPLEMENTATION IN LIBRARIES

<i>Library Grouping</i>	<i>Barrier 1</i>	<i>Barrier 2</i>	<i>Barrier 3</i>	<i>Barrier 4</i>	<i>Barrier 5</i>	<i>Barrier 6</i>	<i>Barrier 7</i>	<i>Barrier 8</i>
All (64 libraries total)	5 (8%)	19 (30%)	32 (50%)	12 (19%)	11 (17%)	12 (19%)	1 (2%)	7 (11%)
Public Libraries (11 libraries total)	4 (36%)	3 (27%)	6 (55%)	0 (0%)	1 (9%)	3 (27%)	0 (0%)	2 (18%)
Successful Libraries (17 libraries total)	2 (12%)	6 (35%)	10 (59%)	1 (6%)	3 (18%)	4 (24%)	0 (0%)	5 (29%)
Problem Libraries (15 libraries total)	0 (0%)	2 (13%)	7 (47%)	5 (33%)	4 (27%)	2 (13%)	0 (0%)	0 (0%)

The above table groups the biggest problems that librarians encountered in introducing and implementing GIS in their library into eight categories. The categories, listed across the top by their number corresponding to that given below, are:

Barrier 1 = Problems with security of the computer environment and/or the data files;

Destruction, inadvertent or otherwise, of the data files;

Barrier 2 = Large amount of staff time required to support the GIS;

Barrier 3 = Large amount of training required for the staff supporting the GIS; complexity of the software;

Barrier 4 = Lack of institutional support and/or lack of a user constituency;

Barrier 5 = Equipment procurement;

Barrier 6 = Procurement of data in format suitable for use by the GIS, either in absolute quantity of the data or in a format readable by the GIS;

Barrier 7 = Unrealistic expectations;

Barrier 8 = No problems encountered; lack of use of the GIS.

Thus it is safe to say that the highest priority for a library contemplating introducing GIS to their patrons is to evaluate the number of staff and the ability of that staff to support the GIS without unduly impacting the rest of the library's users. If this is likely to be a problem, then the library should probably delay introducing GIS to their patrons until the staffing issue can be effectively resolved from a librarywide viewpoint.

Other impediments to GIS implementation differ among the library groupings. Public librarians as a whole identified the issue of computer and data security as being much more important than librarians as a whole (36 percent versus 7 percent for librarians as a whole). This may partly be due to differences in the libraries' clientele. It may also be because research librarians have had to address this issue already in other computer environments.

It is perhaps not surprising that almost one-third of libraries with successful GIS implementations feel that they have not had any significant problems in their projects. Libraries that have had significant problems in implementing GIS identify the major problem as being staffing

issues (covered earlier), procurement of equipment, and lack of support and/or constituency. Problems with hardware procurement led to loss of continuity on the GIS project, decline in staff expertise due to lack of GIS exposure, and loss of user interest. The lack of support (external or internal) or of a clearly-defined constituency leads to inability to adequately plan and promote the GIS implementation and/or the lack of GIS usage in general.

Respondents were also asked to provide advice that they would give to other librarians planning to introduce and implement GIS into a new library. The advice by each respondent was grouped into the eight categories given in Table 2. An individual could provide more than one piece of advice so the totals across a row do not sum to 100 percent.

Without exception, the most frequently given advice was to ensure that the librarians supporting the GIS be provided with enough time and training to learn the software as well as to provide the proper end-user support. Lack of adequate staffing is the second most frequently provided advice by librarians with problems implementing a GIS in their library.

If staffing is adequate, it is also important that the libraries obtain state-of-the-art equipment in terms of hardware. This advice is the third highest ranking one among all library groups following staffing.

Public librarians, more than any other group of librarians, noted that it was important to define the library's constituency group. This is because of the broad range of public library users and the individual public library's inability to serve all of its users all at once. Relevant data sets for users can be developed quickly if a constituency group can be identified so that library efforts can be initially focused on developing the desired data sets for that constituency.

Another significant observation is found in the advice given by libraries that were most successful in implementing Geographic Information Systems. Almost one-half of them advised librarians to implement the GIS in a controlled environment with careful definitions of what should and should not be done by the library. In their opinion, hasty GIS implementations are likely to lead to failures rather than successes. Thus, general conclusions that can be derived from Tables 1 and 2 are:

1. libraries with inadequate staffing levels should not attempt to introduce GIS to their patrons;
2. staff training on the GIS and staff assistance with library patrons is critical to a successful GIS implementation experience;
3. successful implementation of GIS requires an advanced computer configuration which may need to approach state-of-the-art in the micro-computer world;
4. it is important to identify who your user group will be; and

JUE/IMPLEMENTING GIS IN THE PUBLIC LIBRARY ARENA

TABLE 2
ADVICE FOR GIS IMPLEMENTATION TO OTHER LIBRARIES

<i>Library Grouping</i>	<i>Advice 1</i>	<i>Advice 2</i>	<i>Advice 3</i>	<i>Advice 4</i>	<i>Advice 5</i>	<i>Advice 6</i>	<i>Advice 7</i>	<i>Advice 8</i>
All (64 libraries total)	11 (17%)	8 (13%)	13 (20%)	31 (48%)	4 (6%)	6 (9%)	5 (8%)	18 (28%)
Public Libraries (11 libraries total)	0 (0%)	1 (9%)	3 (27%)	4 (36%)	1 (9%)	4 (36%)	1 (9%)	4 (36%)
Successful Libraries (17 libraries total)	0 (0%)	1 (6%)	5 (29%)	10 (59%)	1 (6%)	2 (12%)	1 (6%)	8 (47%)
Problem Libraries (15 libraries total)	4 (27%)	0 (0%)	3 (20%)	7 (47%)	0 (0%)	0 (0%)	0 (0%)	2 (13%)

The above table groups the advice that librarians with experience in implementing GIS in their own library would provide to other librarians into eight categories. The categories, listed across the top by their number corresponding to that given below, are:

Advice 1 = Make sure staffing level is adequate to support the project;

Advice 2 = Develop pre-canned/relevant data sets that are easy for the casual user to access;

Advice 3 = Obtain state-of-the-art equipment, including printer/software protection;

Advice 4 = Provide the supporting librarians with enough time and training to learn the software as well as to provide the proper end-user support;

Advice 5 = Focus on local/regional data sets that provide details of interest to local users;

Advice 6 = Identify constituency;

Advice 7 = Advertise your project through newsletters, GIS bulletin boards, or regular classes and tutorials;

Advice 8 = Implement the GIS in a controlled environment with controlled access; go slow and define the role of the library and librarians carefully.

5. do not rush your efforts to introduce GIS but rather carefully plan the steps that will be taken.

IMPLEMENTING GIS IN PUBLIC LIBRARIES

Model for GIS Implementation in Public Libraries

There have been at least two models that provide guidelines for implementing GIS in a library. Wong (1993) developed three data use models for spatial data in a cartographic library setting. These models were:

1. personal use model, in which each user interacts directly with the GIS where the digital data are stored, leaving out an intermediary such as a librarian. In this model, a librarian has the role of a custodian.
2. chauffeur-driven model, in which users make no direct contact with a GIS and have to rely on a librarian to access and use the digital data. In this model, a librarian assumes the role of a facilitator to cartographic data and the librarian must possess systems and mapping knowledge of GIS as well.

3. adaptive interpersonal use model, which combines the earlier two models into one model. In this model, different levels of services are provided by librarians depending on the sophistication of the user.

Allen (1993) divided the configuration of computers in a library mapping environment into three levels: minimum, medium, and maximum:

1. minimum level—this level requires little expertise from the library staff and places the burden of knowing how to access data on the user. The user will not be able to manipulate data to a great extent and not be able to work with their own data.
2. medium level—this level might consist of a workstation that allows a user to extract data, view, print, and manipulate displayed images.
3. maximum level—this level includes all the functionalities of the medium level plus the capability to import and export data sets from or to other GISs. The overall configuration of the GIS might be better defined as a “GIS laboratory.”

McGraw (1994) analyzed the role of “electronic maps” in public libraries from several major viewpoints (e.g., time, demand, equipment costs). Allen’s article is also referenced and McGraw suggests that the most sophisticated GIS set-up that should be expected in a public library setting is one comparable to the “Medium” level and that even the “Minimum” level may suffice. He suggested that GIS services are probably best housed in larger libraries with extant map collections. Such libraries have comparatively well-trained staffs and should be able to generate adequate demand for the service. He believes that smaller libraries have to be committed financially and politically if they are to succeed, and that network access to GIS could be the best option for small libraries. McGraw identifies several things that public libraries providing GIS services should do:

1. be committed to training staff in GIS techniques and allow staff sufficient time to assist patrons;
2. be willing to enter into partnerships with data producers to make current data available in the library;
3. be committed to maintaining current software and hardware environments; and
4. be developing publicity and training programs for their patrons.

PROPOSED GIS IMPLEMENTATION MODEL FOR PUBLIC LIBRARIES

The two models by Allen and Wong do not explicitly address GIS in the public library environment. The models address the issue from a

depository library/map librarian standpoint. McGraw's paper extended the models into a public library environment, but McGraw does not directly address when the public library setting should be "minimum" or "medium" for GIS. This research paper integrates the models discussed earlier with the data from the survey findings to develop an explicit model for implementing GIS in a variety of public library environments.

The public library environment can be described by its locational setting (e.g., rural branch library, rural central public library, urban branch library). However, there can be so many differences among libraries within a particular locational setting that this breakdown of public libraries by locational setting can be obfuscating. For instance, a rural branch library surrounded by upscale commuters has very different constraints than a similar library branch surrounded by long-time farming residents.

As a result, a better distinction among library branches for GIS implementation purposes is between library branches with a large number of staff members and library branches with only a few staff members. This can sometimes be translated to mean urban and rural library branches, respectively, but this is not necessarily the case. The discussion brought out the importance of adequate staffing as a necessary component for GIS implementation success.

It is perhaps tempting to develop a public library model for GIS implementation based on the staff size and funding level of the public library. But to do so ignores the ultimate reason for implementing the GIS in the first place—i.e., to develop a useful and usable GIS environment for the public library patron. In order to do so, the model must carefully consider the different types and needs of potential GIS users in the library. Allen (1993) identifies three broad spatial data user types:

1. map user who simply wants to make use of an already existing product;
2. personalized map user who wishes to make simple use of existing data to produce a map specifically designed for his or her personal needs;
3. mapmaker, who wishes to acquire, manipulate, and analyze data in a sophisticated manner.

Although Allen does not make the distinction, a necessary one in the GIS environment is between library patrons who are very comfortable in a computer environment—i.e., those that could be called "computer secure"—and those who are uncomfortable in a computer environment—i.e., "computer insecure."

Figure 1 presents a flowchart model for implementing Geographic Information Systems in a public library arena. The flowchart is based primarily upon the most important considerations for successful implementation of a GIS in a library as provided by librarians themselves. These are staffing levels, access to trained staff capable of supporting a GIS,

adequate GIS hardware environment, and an identified user constituency for the GIS. The only additional consideration in the flowchart besides those provided by librarians is the comfort level of the library patrons with computers.

To use the flowchart, a user enters from the "start" box at the top. By answering the questions either "yes" or "no" through the diamond boxes, the user eventually ends up at an outer rectangle that specifies an optimal GIS configuration for a public library given the existing library and user environment. The terms "map user," "personalized map user," and "mapmaker" are from Allen (1993) while the terms "P/U," "A/I/U," and "C/D" correspond to Wong's (1993) personal use, adaptive interpersonal use, and chauffeur-driven models, respectively. The flowchart is iterative. Hence, as the circumstances of the library and/or its patrons change, the optimal GIS configuration will change as well.

Some important points to note from the flowchart include:

1. there is no situation in which a library should try to introduce GIS into its environment if its existing equipment is inadequate;
2. unless the library resides in a market area with a high percentage of computer literate patrons and/or the library has ready access to computer-literate staff, Allen's map user option is probably the only GIS implementation option that should be considered;
3. five of the ten possible paths through this flowchart lead to a map user option as the best choice, while only two of the possible ten lead to a mapmaker option.

Thus, a good starting assumption for public librarians introducing GIS into their library would be that the library should support only map users until other circumstances suggest otherwise. Only after a period of public exposure to the simpler forms of digital spatial data sets should a public librarian contemplate supporting a mapmaker option and then only if data sets are readily available in addition to a computer literate staff.

Another way to implement GIS in a public library setting is provided in Table 3. This table analyzes the library environment relative to two important variables, that of computer "comfortableness" of the library patron and library staffing. There are four possible combinations of the two variables, and Table 3 lists all four of them in the left-most column. Given the staffing availability of a particular library branch, the middle column gives the resulting computer use model possibilities developed by Wong (1993). For instance, a library branch with a small number of staff can best support Wong's personal use data model in which each user interacts directly with the computer and any GIS software because a public librarian would most likely not have a large amount of time to handle detailed and involved questions by a single patron.

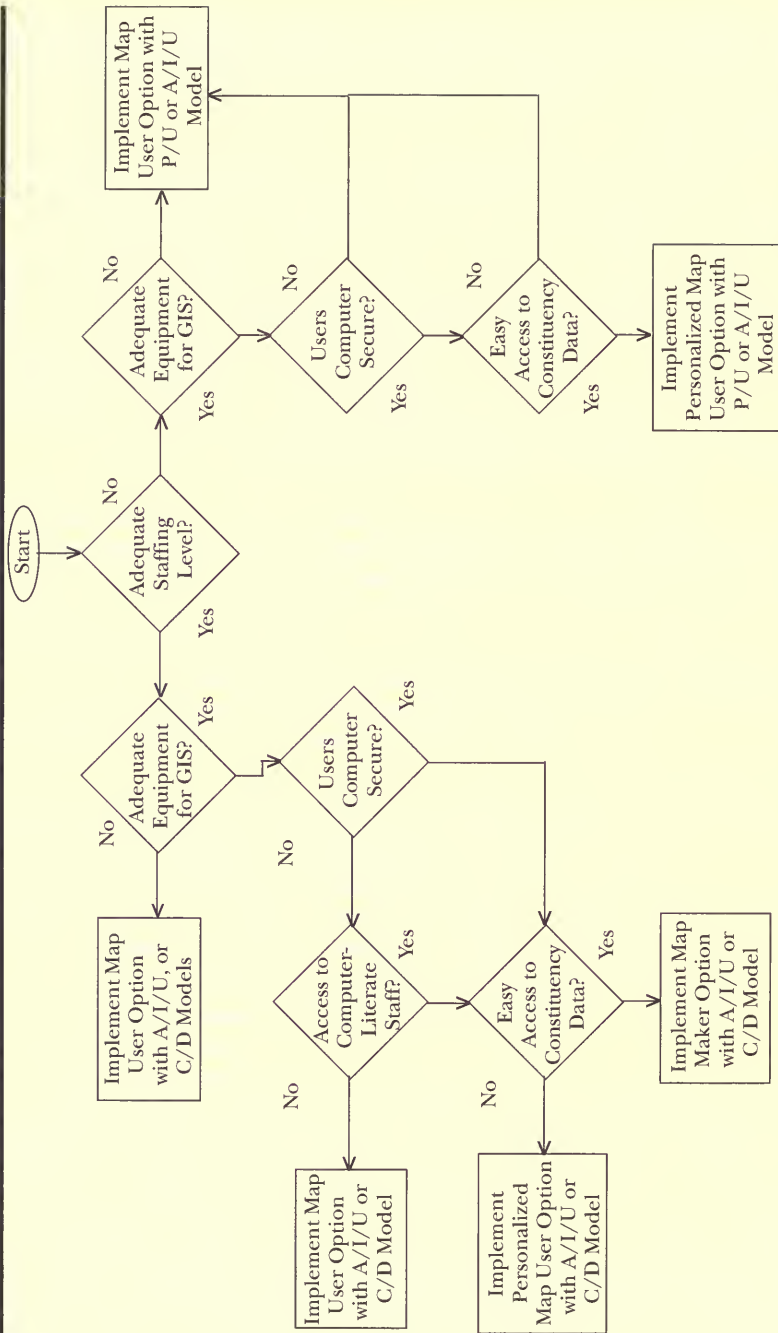


Figure 1. Decision flowchart for implementing GIS in libraries

The final column of Table 3 lists Allen's three types of spatial data users within each of the four library types relative to the optimal computer use model. For instance, a map user that is uncomfortable with computers could probably still be supported in a public library environment with a limited number of library staff if he or she is merely a map user. This is because of the relative simplicity of such software. The same "computer-insecure" user may not be able to be easily supported in such a public library environment if he or she wanted a personalized map, however, because of the limitations of library staff, that user should probably go to a better staffed library (library type 2 and 4 of Table 3).

Based on their comments about GIS software complexity from the majority of librarians and the lengthy time requirements to assist library patrons in their GIS usage, it can probably be argued that library type numbers 3 and 4 are very scarce at this point in time. The reason is the lack of exposure of most casual library patrons to important GIS concepts at this time; even "computer secure" library patrons essentially become "computer insecure" in a GIS environment.

If this assertion is true, then one would expect that "computer secure" individuals would develop with adequate training and exposure to GIS. The environment in which this would most likely be found are in the academic and research libraries participating in the ARL GIS Literacy Project. In fact, the ARL participants with the most satisfying experiences and the most usage of GIS are in a library type number 4 environment. They usually have dedicated individuals (or students) that are staffing the library as part of their own training and are assisting others in learning GIS. The library patrons in this case are often students who are already comfortable with computers and probably only need some minimal exposure and training to GIS concepts before they become "computer secure" enough to produce their own digital maps.

Westcott (1994) identified some of the major issues to be addressed in implementing a GIS in a public arena setting, of which the public library is obviously one:

1. determining the market;
2. developing desirable data, products, and services;
 - a) What data do you have?
 - b) How do you market your data?
 - c) How do you satisfy technical demands?
 - d) What services do you provide?

The model presented in Figure 1 addresses all of these issues in some manner. The market is determined during the evaluation of the computer comfort level of the average library patron. The identification of

TABLE 3
OPTIMAL COMPUTER USER MODEL FOR A PUBLIC LIBRARY BRANCH

<i>Library Type Number</i>	<i>Library Branch Characteristics</i>	<i>Optimal Computer User Model (from Wong, 1993)</i>	<i>Digital Data User Types (from Allen 1993)</i>	<i>Ability of Library to Support Allen's Data User Type</i>
1	"Computer Insecure" Library Patrons, Small Library Staff	Personal Use Model	Map User Personalized User Map Maker	Good Poor Poor
2	"Computer Insecure" Library Patrons, Well-staffed Library	Adaptive Personal or Chauffeur-driven Use Model	Map User Personalized User Map Maker	Good Good Depends
3	"Computer Secure" Library Patrons, Small Library Staff	Personal Use Model	Map User Personalized User Map Maker	Good Depends Depends
4	"Computer Secure" Library Patrons, Well-staffed Library	Personal Use or Adaptive Personal Use Model	Map User Personalized User Map Maker	Good Good Good

The above table shows the optimal computer user model for a public library branch given the computer literacy of its library patrons and the library staffing size. The ability of each of the four types of public library branches to support the three different types of digital data user types (from Allen, 1993) given the optimal computer user model for each type of library branch is given in the right-most column.

the constituency group (implicit in the flowchart model) will require determining what data a library may have and how to market it. The determination of the mapping option for the library patron will determine the technical demands and the level of service that will be provided.

THE PUBLIC LIBRARY USERS OF GIS (PLUS+GIS) PROJECT

The PLUS+GIS project is a cooperative effort of Florida State University researchers, urban and rural public libraries in Florida, and private companies. The research objectives include:

1. developing a testbed for introducing GIS into rural and urban public libraries;
2. making digital spatial data accessible and usable for public library patrons; and
3. identifying barriers to accessing digital spatial data for public library users and developing methods to overcome those barriers.

This project has workstations set up at public libraries in four Florida counties: Leon, Franklin, Jefferson, and Wakulla. In addition to the pub-

lic domain Landview software, ESRI has donated copies of ArcView for this project.

Major tasks that are being conducted to address the ability of public library users to find the data they need include: (1) spatial metadata standards analysis and evaluation relative to library cataloging practices; and (2) development of surveys of casual user spatial information needs.

Two research tasks that directly improve the ability of casual users to utilize GIS software include: (1) user interface design and software development/modifications using ESRI's AVENUE product; and (2) training material evaluation and development.

This project will review existing GIS interfaces to evaluate their usability by casual users. Based on this evaluation, the researchers will, in an iterative process, develop, test, evaluate, and modify alternative GIS designs and applications in an attempt to determine the elements and components of a user-friendly GIS interface in a public library setting. Examples include more restrictive or "pre-canned" GIS and more intuitive and self-training user interfaces and hypermedia in providing access to GIS data.

Although the PLUS+GIS project is research-oriented, it does include actual placement of GIS and digital spatial data sets into a public library setting. PLUS+GIS addresses the major GIS implementational issues for libraries derived from Tables 1 and 2 by taking the following steps:

1. Library staffing is supplemented with trained geography undergraduate students from Florida State University. These students were volunteering in the Map Resource Center (MRC) in the Leon County Public Library. The MRC is a collection of hard-copy maps accessible to the general public.
2. Training for library staff as well as to the geography students is provided on a regular basis by the staff of the Spatial Analysis, Research, and Training (SART) Program at FSU. SART staff also visits the libraries regularly to evaluate the process of the study and the implementation of the GIS in each of the libraries.
3. All of the participating libraries have a state-of-the-art computer configuration as of early 1995.
4. The user constituency groups have been identified to be local citizens who are interested in accessing local planning data. The Leon County Growth and Environmental Management Office is actively working with the FSU team to place their spatial data sets on a GIS workstation in the local public libraries.
5. The introduction of the GIS to the library patrons has been carefully planned. Because the placement of the GIS in each of these public libraries is part of a larger research project, a careful evaluation of

the steps and procedures being undertaken is occurring. In addition, there is a one-month "break-in" period of the GIS for the public librarians prior to the official announcement of the GIS software to the general public.

CONCLUSION

The introduction of GIS into the public arena has been met with both success and failure in libraries. With the decreasing costs of hardware, costs are disappearing as a major barrier to providing this type of service to library patrons. The issue is more whether the library should provide such a service and, if so, exactly how much is adequate.

This paper identifies the major problems that have been encountered by some of the pioneering libraries in this arena, as well as some of the advice that such libraries would provide to others contemplating similar actions. From this information, a flowchart was developed to assist librarians in making a decision for their own particular situation.

Part of the problem with the introduction of GIS into a public library setting is that there has been very little research on ways to make accessing and use of digital spatial data easier for the casual data users, a category into which most public library users would fall. The PLUS+GIS project housed at Florida State University addresses this need.

APPENDIX A

Library Name: _____

Respondent's Name: _____

1. Please describe the computing environment in which your (ARL) GIS library project resides—i.e., is it networked, stand-alone, etc.?
2. Please estimate the relative percentages of your clientele in your GIS library project.
 PUBLIC _____% SCIENTISTS/RESEARCHERS _____%
 BUSINESS _____% ORGANIZATIONS _____%
 STUDENTS _____% OTHERS: _____%
3. How long has your library had GIS software in the library for your client's use?
4. Which GIS software packages do you have?
5. What other type of map-making/displaying type software (e.g., Landview, Street Atlas) do you have within your facilities?
6. How often is the GIS used during an average day? _____
 By how many individuals? _____
7. What digital data sets are you making accessible to your GIS users in the library?
8. What is the most frequently-used data set?
9. Is there a most frequently-requested data set that you do NOT have available?
10. Is the GIS workstation connected to the Internet? Yes No
11. Do you use or are planning to use the GIS for library management or library planning?
12. What is the amount and type of training that is provided to librarians who support the GIS system for your clientele? From whom was the training material obtained?
 What is the amount and type of training that is provided to your clientele who may wish to use the GIS? From whom was the training material obtained?
13. Who provides the technical support for the GIS hardware/software?
14. What has been the greatest problem with the GIS implementation in your facilities?
15. What would be your advice to other libraries that are introducing GIS to their clientele?
16. (For ARL Libraries Only) Are you supporting or providing any technical assistance to any public libraries as part of your project?
17. Do you know of any other libraries, especially public libraries, that are using GIS?

REFERENCES

- Allen, R. S. (1993). Configuration of computers in map libraries. *Special Libraries Association Bulletin*, 173, 15-23.
- Huxhold, W. E. (1991). *An introduction to urban geographic information systems*. New York: Oxford University Press.
- Kofron, C., & Watts, A. (1993). St. Louis Library disseminates public information. *Geo Info Systems*, 3, 46-50.
- Kollen, C., & Baldwin, C. (1993). Automation and map librarianship: Three Issues. *Special Libraries Association Bulletin*, 173, 24-38.
- McGlamery, P., & Lamont, M. (1994). Geographic information systems in libraries. *Data-base*, 17, 35-44.
- McGraw, R. (1994). *Electronic maps in public libraries*. Unpublished practicum at the Cleveland Public Library.
- Westcott, B. (1994). Meet the challenge of public spatial data demand. *Geo Info Systems*, 4, 56-59.
- Wong, M. (1993). Exploring the impact of digital cartographic data on map librarianship using data use models. *Special Libraries Association Bulletin*, 173, 2-14.

The St. Louis Public Library's Electronic Atlas: A Successful GIS Application in the Public Library Environment

In 1974, the U.S. Census Bureau published a collection of thematic maps from the 1970 census. The maps found a wide audience during the decade. However, the bureau did not produce a companion set from either the 1980 or 1990 censuses. Other organizations have responded to the demands of data users with comparable products. Rather than producing a published set of thematic maps, the St. Louis Public Library (SLPL), with the Illinois State Data Center at Southern Illinois University at Edwardsville, decided to produce an electronic set of thematic maps using 1990 U.S. Census data. The St. Louis Public Library's electronic atlas began with clearly defined goals, modest expenditures, and a wide range of community partnerships. As the electronic atlas enters its third year of operation, it is possible for the participants to discuss the components of a successful implementation of Geographic Information System (GIS) technologies in a public library environment.

The St. Louis Public Library (SLPL) dates its beginning to the Public School Library Society of St. Louis, which was established on February 3, 1865. Shortly after its founding in 1865, the library was designated a federal depository library, making St. Louis Public Library the oldest such depository west of the Mississippi River. The library's federal document collection is one of the oldest and largest in the region, numbering approximately 1.5 million items including approximately 110,000 paper maps.

As part of its mission, the library has actively worked within the St. Louis community to see that the public is made aware of the riches in its collections. The library pioneered in the utilization of electronic databases by the public, placing its catalog in high schools, public agencies, and other community facilities across the city and encouraging public dial-in to the library's catalog through desktop. The library provides public access computers to both children and adults for activities ranging from word processing to full-text and bibliographic database searching. In addition, the library has actively worked within the St. Louis community to see that the public is made aware of the resources in its local, state, and federal document collections by conducting community workshops, training sessions, and participating in the data user community.

At the same time, the Illinois State Data Center Cooperative program, located in Regional Research and Development Services (RRDS) at Southern Illinois University-Edwardsville, has aggressively pursued a data dissemination partnership arrangement with the St. Louis Public Library and other libraries in downstate Illinois and eastern Missouri. Both institutions are active participants in the St. Louis metropolitan data user community with memberships in such groups as the United Way Research Committee; the American Statistical Association, St. Louis Chapter; and Midwest Gateway Chapter of the Urban and Regional Information Systems Association (URISA).

The St. Louis Public Library's Decennial Census collection dates from 1790. Many of these earlier publications continue to be important reference sources years after their original publication. One such publication followed the 1970 census when the U.S. Census Bureau produced a series of thematic maps using popular census data items. The maps were published under the title, *Urban Atlas, Tract Data for Standard Metropolitan Statistical Areas: St. Louis, Missouri—Illinois* (U.S. Bureau of the Census, 1974) and contained topical colored maps at the census tract level. The maps classed and displayed data reflecting a range of themes of interest to urban demographers. These maps were used in a variety of census data applications to illustrate the spatial distribution of special population groups, trends in population change, and interrelationships among classes of census variables. Typically, library users would photocopy the maps and incorporate the graphics in reports and presentations. The colors used in the maps were sufficiently different to allow black and white photocopying and reduction.

The urban atlas was a very successful product. The collection of thematic maps allowed users to illustrate data and conceptual points graphically, and the tract reference maps provided a handy resource for identifying geographic area locations by census tract number. Unfortunately, the Census Bureau did not produce companion sets of maps following the 1980 and 1990 censuses. However, because of its popularity, ease of use, and ease of comprehension, other users and disseminators of census data thought it important to undertake the task of producing an atlas product. The East-West Gateway Coordinating Council (EWGCC), which serves as the major transportation planning agency for the St. Louis metropolitan area, produced a two volume set of maps following the 1980 census which went beyond the data elements and urban geographical components of the bureau's 1970 urban atlas (EWGCC, 1982). EWGCC's urban atlas extended mapping coverage to the less urbanized counties of the metropolitan area. The themes depicted in the maps were also expanded.

For 1990, the St. Louis Public Library and Regional Research and Development Services at Southern Illinois University at Edwardsville planned to produce an urban atlas which would replicate many of the

same maps found in the bureau's 1970 urban atlas and East-West Gateway's 1980 urban atlas.

As the SLPL and RRDS began to work on the 1990 version of an urban atlas, several questions and concerns emerged from the discussions. First and foremost was the cost of a paper product. Second, questions arose about the classing of the census variables. With the availability of Geographic Information System technology, it became easier to experiment with a wider range of classing schemes, and the system developers anticipated that users would want to utilize these options in the preparation of their maps. Third, the amount of data that would be mapped in a paper atlas was limited to a few tables from the complete count and sample tape products.

With the intent of providing as much data as possible, it became obvious that there would be requests for other thematic maps and, ultimately, the additional data compilation would be more beneficial to users than only maps by themselves. For these reasons, the development of a paper 1990 urban atlas became less of a priority and, indeed, took a back seat to data processing and planning operations for the release of tapes and CD-ROM census products following the 1990 census.

A number of significant developments took place in the Geographic Information System software market during 1992. The cost of GIS software became less expensive, and the capabilities in GIS toolboxes were increasing. Several software developers choose to segment their GIS software markets by offering desktop mapping products to be used with their GIS toolboxes. One of these companies was Environmental Systems Research Institute (ESRI). With the availability of tract maps in digital form, census data in machine readable form, and software to display both forms of data, the idea of generating an "electronic atlas" began to take shape.

The St. Louis Public Library had planned a large map exhibit highlighting maps from its own collections for fall 1992. As part of the "Roads, Rails, Rivers, and Rifles" exhibit, the library wanted to look at the possibilities of electronic mapping. Plans for the exhibit called for workstations that provided the public with an opportunity to work with several commercial software atlases. The *St. Louis Public Library Electronic Atlas* was to be the highlight of this display.

In July 1992, a formal proposal was made to create an electronic atlas of census tract maps and data from the complete count and sample compilations. The electronic atlas has been limited to St. Louis City and County census tract data and maps. As the original concept was developed, the electronic atlas was to have thematic displays of census tract data allowing for the addition of new levels of geography or data over time. The budget required that we minimize system development costs, assuring a more cost-effective solution than hard copy color maps.

One additional requirement of the library system was that it needed to be both user friendly for the novice computer user while offering an environment appropriate for those more experienced in using a GIS. Unlike most other GIS workstations, this one was to be in a public area and would be accessible on a walk-up basis to library patrons of all ages and skill levels.

To implement an electronic atlas for general public use, the St. Louis Public Library purchased an IBM-compatible computer with the following features: an 80486-50mhz CPU with 8Mbytes of random access memory, a 211Mbyte hard disk drive, two floppy drives, one parallel and two serial ports, a tower case, a keyboard, a mouse, a 15 inch 1024 x 768 noninterlaced monitor, and a Windows accelerator video board. In addition to the hardware, SLPL purchased DOS 5.0, Microsoft Windows 3.1, ArcView (ESRI), and two Arc/Info digital map bases. This configuration was bundled as an ArcView Turnkey System by RRDS-SIUE with software and data installation, delivery, and two hours of training.

The library selected the street maps and address databases for St. Louis City and St. Louis County as their base maps. Although the St. Louis Public Library district is the same as the city of St. Louis, many data needs overlap political boundaries. The Regional Commerce and Growth Association of St. Louis (RCGA) had digital census tract maps and street maps prepared by RRDS-SIUE using Arc/Info and the same TIGER data. The digital data were made available to the library gratis. This provided the capability of overplotting census tract outline maps on top of the street base maps.

The ArcView software contains a special capability for building and saving maps or thematic "views" of tabular data. With this capability, thematic maps were created and stored as data files. The thematic maps are available to the user through program groups and program items in the Windows environment. The series of census data views are organized by city and county. The view and data dictionaries are organized in a separate program icon grouping, assuring ready access to needed documentation. In addition to views of census data, other geographic data products have been incorporated. With ArcView's ability to display satellite imagery, a spectrally enhanced Landsat Thematic Mapper image and a SPOT Image Corporation panchromatic image of a portion of the St. Louis metropolitan area were included.

Initially, access to the full range of functions available on the system was unlimited. The library staff soon determined that some limitation of patron access to system functions was essential—e.g., limiting the ability to change predefined maps/views, changing tabular data, and changing some of the system defaults through access to DOS. The pull-down menu options on the menu bar (just below the Program Manager banner) have been al-

tered to restrict user access to DOS and the Windows environment. Users cannot alter predefined views, spatial data, or tabular data. These changes were made incrementally over the first two months that the system was in place. The present limitations seem adequate to both protect the system and provide users with the full range of options necessary to produce the types of maps needed. Such limitations are essential for user and general system success in a public environment such as the library's.

Using the system is relatively straightforward. By double clicking on a program group, program icons become available to the user, the ArcView program is executed, and the corresponding view is automatically brought to the screen. With the street base maps and the address indexes for both St. Louis City and County, the electronic atlas provides users with the capability of performing address location queries and creating point maps from address data sets. By choosing the Street and Address Index icon, the system makes the address index available. Selecting the mailbox icon in the graphics window and typing in an address will generate a screen with an arrow pointing to the block of the address requested in the query.

Although reference mapping and address location queries are important features of the system, the main purpose is to give users access to census data in a way that is easily understandable and supportive of additional database querying. This is the system's strong suit. Using the identification function in the toolbox, a user can move the arrow to any census tract polygon on the screen and view the corresponding database record. Also, it is possible to view the entire database while a particular view is displayed.

The system is also able to overplot tract boundaries and select records using spatial criteria or by simply pointing to a feature in the graphics window. Taking this a step further, the electronic atlas can be used to perform simple spatial analyses. For example, a user may wish to determine the total population within a one mile radius from a given address. The electronic atlas could be used to identify the address, overplot census tracts, and then select tracts within one mile of the address. Statistics would then be generated on total population from selected tracts within the mile.

An extremely powerful element of ArcView and the electronic atlas is the capability to display satellite data of portions of St. Louis City and County from classified SPOT and Landsat imagery. With rectified imagery, registered to the same ground-based coordinate system as the census tract and street base maps, it is possible to plot the tract or street maps on top of the imagery.

During the first few months of operation, the system was used largely by patrons from the not-for-profit community, students working in the health care field, and small business developers working on business plans.

Many patrons have required staff assistance working with the system. Others, however, have been able to proceed quite independently. If patrons understand Windows, they can learn to use the system quite easily; those without computer and Windows experience have a more difficult time.

Staff help new users get started and oriented to census geography and the concept of a Geographic Information System. Very few patrons of the library are aware of GISs and the potential of the electronic atlas for assisting them in answering questions. Many have returned after an initial successful experience. Staff assist patrons by suggesting views and legend classifications. For branch patrons and those who cannot or do not want to operate the system, we prepare maps. When users have more complicated data requests or require more functionality than we are able to offer, they are referred to the Illinois or Missouri State Data Centers. They, in turn, refer users to the library when our system will meet their needs.

The library began the project with limited and well-defined goals. The focus of the system was to provide census maps and data that would facilitate the various types of analyses and uses of census tract data. The prearranged thematic views of the more commonly used data items are a recognized limitation. However, offering system users the ability to change classing schemes has compensated for some of these limitations. The developers have also realized that storage and data access was a finite resource. Not all spatial data could be offered on the system with the immediate access that we sought. Because we were able to leverage resources, obtaining data gratis from other community information sources, especially the Regional Commerce and Growth Association, we were able to minimize library costs.

Several other internal organizational issues played a part in the success of the project. The electronic atlas was developed jointly by the library staff who know about user requests, and the State Data Center consultant who knows about GIS and ArcView. In addition, the project had an internal organization champion who was a principal in all phases of project development. Keeping the system open to the public was endorsed strongly by the professional GIS community and the library administration. The library administration was also kept informed during all stages of project development. They were asked about the project when away from the library and were proactive in thinking about the electronic atlas when developing information to share about library services. Implementation of the electronic atlas produced software technologies which were useful to other open-access data systems in the library. Finally, and critical to the ongoing success of the project, the library has received support from various professional communities and library constituencies who have maintained interest in the project.

The library views the electronic atlas as only one component of a larger cartographic database. The staff continues to explore the possibility of using the maps available on ArcUSA for indexing portions of the library's map collections. An in-house index to the paper map collection using askSam has been prepared. The electronic atlas has become a natural extension of St. Louis Public Library's electronic information dissemination program. ArcView has enabled the St. Louis Public Library to "create" a reference tool, much as we would have purchased a reference book in earlier, less interesting, times.

REFERENCES

- East-West Gateway Coordinating Council. (1982). *An urban atlas for the St. Louis Metropolitan Area*. St. Louis, MO: East-West Gateway Coordinating Council.
- U.S. Bureau of the Census. (1974). *Urban atlas, tract data for standard metropolitan statistical areas: St. Louis, Missouri—Illinois*. Washington, DC: USGPO.

MARK JOSELYN* & SHERYL G. OLIVER†

Digital Spatial Data of Illinois on CD-ROM, Department of Energy and Natural Resources

INTRODUCTION

The Illinois Department of Energy and Natural Resources (ENR) and its divisions, the Illinois State Geological Survey, Illinois Natural History Survey, Illinois State Water Survey, Hazardous Waste Research and Information Center, Office of Research and Planning, and Illinois State Museum, have been actively involved in the use and application of GIS technology for over ten years. A significant result has been the development of one of the most comprehensive spatial digital databases, at the state level, anywhere in the country. In order to make this information more accessible, many of ENR's statewide data sets were published on CD-ROM in April 1994.

This publication is the first of its kind in the country from a state perspective. It signifies technological advancement, but it also reflects the evolution of an institutional attitude toward data, specifically that data of this type are a resource to be shared and utilized by citizens and groups with an interest and concern for Illinois, its people, and its resources. This publication is intended to make this valuable and dynamic resource more accessible and available to other state agencies, libraries, schools, public interest groups, and the private sector. An anticipated effect will be to increase knowledge and consideration of geography relative to the state's people, resources, and related issues. Virtually all aspects of the built and natural environments in which we live have spatial components. The data contained on this CD-ROM provide a basis for integrating these spatial components into research, education, and management.

DEVELOPMENT

Data sets contained on the CD-ROM were primarily constructed on a project by project basis in response to programmatic needs of the divisions of ENR. Others—particularly infrastructure, hydrography and census

*This article was reprinted with permission from *Illinois GIS and Mapnotes*, Volume 12, Number 1, Fall 1994. Publications Department, Center for Governmental Studies, Northern Illinois University, DeKalb, Illinois, 60115.

†Presenter and evaluator of the paper at the conference.

tracts—were derived from U.S. Geological Survey or U.S. Bureau of the Census data files. While some modifications were made specifically for this publication, none of the data were produced simply for inclusion on the CD-ROM. Rather, they have been used extensively for more than ten years by all of ENR's divisions for basic research and in response to project needs. The data contained on the CD-ROM represent thousands of hours of time and effort when all aspects of data collection, validation, and translation are considered. Most of the information is at a scale of 1:100,000 although scale varies depending on the map layer. Data at a scale of 1:100,000 are suitable for regional and local, but not site specific analysis and use. National map accuracy standards equate to a positional error of 167 feet at this scale.

CONTENTS

The CD-ROM contains the following map layers:

- County*
 - Archaeologic probability
 - Airports
 - Census tracts
 - Flood zones
 - Hydrography
 - Incorporated areas
 - Natural areas inventory
 - Nature preserves
 - Railroads
 - Roads
 - Utilities
 - Wells
- Statewide*
 - County boundaries*
 - Interstate, State, US Highways
- House Districts*
- Landfills*
- Public Water Supply Intakes*
- Quadrangle boundaries*
 - 7.5 x 7.5 minute series
 - 30 minute x 60 minute series
 - 1 degree x 2 degree series

* These data are duplicated by county

*Senate Districts**Towns***Townships**

Data sets of particular value and interest, all of which were developed entirely by one of ENR's divisions, are discussed in more detail below.

The archaeological probability map (1:250,000) was developed by the Illinois State Museum. It depicts areas believed to possess a high archaeological potential and is intended for coarse regional analysis. It has assisted agencies in responding to requirements of the Illinois State Agency Historic Preservation Act (P.A. 86-707) which mandates state government to preserve, restore, and maintain the historic resources of the state.

The wells coverage, developed and continuously updated by the Illinois State Geological Survey, is the most comprehensive database of its kind in the country. At the time of publication, it contained 335,695 individual well and boring locations throughout the state, their well type, and a unique identification number used to relate back to over sixty fields of attribute data recorded for each location. The accuracy of location for most features is considered to be +/- 100 feet.

The Illinois Natural History Survey created the map layer of Illinois Natural Areas (1:24,000) and Illinois Nature Preserves (1:24,000) as identified by the Illinois Department of Conservation. These areas represent some of the finest of Illinois' natural resources and include areas of unique biologic, geologic or natural value.

The Illinois State Water Survey developed the unincorporated flood zone boundaries (1:12,000, nominal) data set for all of Illinois. Federal Insurance Rate Maps (FIRM) and Federal Emergency Management Agency (FEMA) maps were the primary source for this map layer. The solid waste landfill sites active in 1993 represent a subset of a much larger database, the Inventory of Landbased Disposal Sites, maintained by the Hazardous Waste Research and Information Center.

The 1992 Census data provide a comprehensive assessment of basic demography and ethnicity at the census tract level for the entire state. The Office of Research and Planning also contributed the legislative boundaries for State Senate and House districts as of 1992.

The roads, railroads, hydrography, and utilities coverages were derived from USGS Digital Line Graph files. Significant value was added by recoding the attribute values to facilitate use. The roads and railroads coverages have both been updated. All major highways, as of 1993, were added to the roads coverage while railroads were updated as of 1991 and

* These data are duplicated by county

include active/abandoned status and ownership. Documentation, based on the Federal Geographic Data Committee Spatial Metadata Standard (draft of January 25, 1994) is included on the CD-ROM for each layer.

LICENSING AND LIABILITY

The data, with the exception of the wells coverage, are not explicitly copyrighted. However, the licensing agreement request states that data are not to be redistributed without prior knowledge of ENR. Liability concerns are addressed in the licensing agreement which states, in part, “the burden for determining fitness for use lies entirely with the user.”

FORMAT AND AVAILABILITY

Data included on the CD-ROM were produced using ARC/INFO from ESRI, Inc., Redlands, California, and are all in ARC/INFO format. This does not indicate an endorsement of this product but simply reflects the primary software currently used by ENR and its divisions. ARC/INFO, ArcView, or ArcView2 software on a UNIX workstation with a CD-ROM drive are required to access the map information. The CD-ROM also works on Macintosh computers running ArcView2. While this does not address the needs of all potential users, it does provide a “plug-and-play” product that can be used immediately, assuming proper hardware and software environments. A copy of the CD-ROM is available, free of charge, to government agencies, schools, and libraries, and at a charge of \$100 to the private sector and private individuals.

ENR is dedicated to the maintenance and improvement of these data and welcomes any comments about their content, format, or utility. People interested in additional information should contact Sheryl Oliver, GIS Manager, Office of Research and Planning, 325 West Adams Street, Springfield, Illinois 62704 (217) 785-7575.

Contributors

BARBARA P. BUTTENFIELD is Associate Professor of Geography, SUNY—Buffalo and a Research Scientist at the National Center for Geographic Information and Analysis (NCGIA). Her research interests include map generalization, cartographic knowledge representation, and roles for map animation and hypermedia in GIS. She has led three NCGIA Research Initiatives including Multiple Representations, Visualization of Spatial Data Quality, and Formalizing Cartographic Knowledge. She is one of the co-principal investigators on the Alexandria Project, funded by NSF, ARPA, and NASA, with the goal of implementing and evaluating a digital library for digital spatial databases distributed across local and wide area networks. She is President of the American Cartographic Association, Associate Editor of the *Journal of Cartography and GIS*, and a member of the Mapping Sciences Committee at the National Academy of Science.

ROBERT LEE CHARTRAND has served as Senior Specialist and then Senior Fellow in Information Policy and Technology for the Congressional Research Service of the Library of Congress. For more than 25 years, he has been an advisor to members and committees of the U.S. Congress, preparing major studies, conducting special seminars, and holding consultations on the roles and applications of information and its technologies within government and society. In addition to authoring the major congressional study, *Information Technology for Emergency Management*, and editing, with K.C. Chartrand, the recent award-winning book, *Strategies and Systems for Disaster Survival*, Mr. Chartrand has written articles for *The Futurist*, *The ASIS Bulletin*, *Hazard Monthly*, *Natural Hazards Observer*, and *CRS Review*. He is the recipient of the Life Achievement Award at the International Emergency Management and Engineering Conference and received the 1985 ASIS Award of Merit and the Library of Congress Award for Superior Service in 1988.

MICHAEL A. DOMARATZ is a member of the secretariat for the Federal Geographic Data Committee (FGDC). He works with the committee to

CONTRIBUTORS

develop goals and helps to design and carry out strategies to attain these goals. His current work concentrates on the implementation of a framework of data for the National Spatial Data Infrastructure. He also provides information to federal officials and agencies, state and local governments, the private sector, and academia about the FGDC's activities. His previous assignments to the committee include participating in the development of a metadata standard and data clearinghouse and serving as the FGDC's executive secretary. Mr. Domaratz is a cartographer in the National Mapping Division of the U.S. Geological Survey (USGS). Since joining the USGS in 1978, he has worked in the areas of spatial data collection, computer-assisted mapping, geographic information systems, and spatial data structures.

MYKE GLUCK is an Assistant Professor, School of Library and Information Studies, Florida State University and is affiliated with FREAC (Florida Resources and Environmental Analysis Center), Florida State University in Tallahassee. He conducts research on the usability, usefulness, and visualization of geospatial data and geographic information systems. He has published several articles in these areas including an upcoming article in the May issue of *JASIS* on competence in information systems using a spatial data platform. Mr. Gluck is co-principal investigator for a grant from the Federal Geographic Data Committee addressing citizen access to the spatial data clearinghouse currently under construction. This is part of the larger PLUS+GIS project at Florida State University partially sponsored by ESRI, Inc. Mr. Gluck has been a frequent presenter on geospatial information topics at ASIS meetings and to a range of other organizations including the Special Library Association, the World Computer Graphics Society, and the Association of American Cartographers.

LINDA L. HILL is a senior research scientist, College of Library and Information Services at the University of Maryland College Park. She provides coordination between the library/information science and services community and the NASA GCDIS and Digital Libraries programs. Dr. Hill's career has focused on the direction and development of library/information services and systems and research, including consulting, federal government contracting, abstracting and indexing, and public library services for the business and scientific communities. Her research has centered on the effectiveness of concept representation for online information retrieval, particularly geographic concepts and user behavior in information seeking activities. Dr. Hill was most recently a senior analyst with Information International Associates in Oak Ridge.

MARK G. JOSELYN is an Assistant Supportive Scientist at the Illinois Natural History Survey at the University of Illinois at Urbana-Champaign.

CONTRIBUTORS

DEAN K. JUE is currently at the Florida Resources and Environmental Analysis Center at Florida State University. His responsibilities include technical assistance to state agencies and counties especially as related to Geographic Information Systems (GIS). He currently is working with Christie Koontz on developing techniques and methodologies for introducing GIS into a public library environment. He received the Federal Geographic Data Committee's National Spatial Data Infrastructure grant in 1994.

CHRISTIE KOONTZ is an Assistant in Research, Florida Resources and Environmental Analysis Center, Florida State University, Tallahassee. She has presented papers at American Library Association Conferences, at the Digital Libraries '94 Conference, at the 1994 Conference on Elementary-Secondary Education Management Information Systems, and a conference sponsored by the U.S. Department of Education National Center for Education Statistics. She has authored several publications including the forthcoming book, *Library Facility Siting and Location Handbook* (Westport, CT: Greenwood Publishing Group, Inc.). Ms. Koontz was awarded the Carroll Preston Baber Research Award in 1991/1992 and a Georgia Endowment for the Humanities grant for a videotape presentation which documented the history of southwest Georgia that the sponsoring regional library served. She is a member of the Public Library Association, Association for Library Services to Children, the Library Administration and Management Association, the Intellectual Freedom Roundtable, and the Library Research Roundtable.

MARYLYNETTE LARSGAARD is Assistant Head, Map and Imagery Laboratory of the Davidson Library at the University of California at Santa Barbara. Ms. Larsgaard has worked in map librarianship since 1969. Her publications include: *Map Librarianship: An Introduction* (2d ed., Littleton, CO: Libraries Unlimited); *Topographic Mapping of the Americas, Australia, and New Zealand* (Littleton, CO: Libraries Unlimited); and *Topographic Mapping of Africa, Antarctica, and Eurasia* (Santa Cruz, CA: Western Association of Map Libraries). She also served as Executive Editor of the *Information Bulletin* (Western Association of Map Libraries) from 1989 to 1994.

RAY R. LARSON specializes in the design and evaluation of information systems and the evaluation of user interaction with those systems. He has worked as a programmer/analyst with the University of California (UC) Division of Library Automation (DLA) where he was involved in the design, development, and performance evaluation of the UC public access online union catalog (MELVYL). His research has concentrated on the design and evaluation of information retrieval systems with an emphasis on online library catalogs. Professor Larson was a faculty investigator on the Sequoia 2000 project where he was involved in the design and

CONTRIBUTORS

evaluation of a very large-scale network-based information system to support the information needs of scientists studying global change. He is a faculty investigator on the UC Berkeley Environmental Digital Library Project. Professor Larson is also principal investigator for the "Cheshire Demonstration and Evaluation Project," a next-generation online catalog and full-text retrieval system sponsored by the U.S. Department of Education.

WILLIAM E. MOEN is a Ph.D. candidate at the School of Information Studies, Syracuse University. His dissertation is titled, "Understanding the Information Technology Standards Development Process: Towards a Conceptual Model Based on a Case Study of ANSI/NISO Z39.50." He was project coordinator and co-principal investigator on the research project that resulted in the Government Information Locator Service (GILS). He has published articles and reports on Federal information policies, on organizing networked resources, and on electronic networking and libraries, including a book co-authored with Charles McClure and Joe Ryan entitled *Libraries and the Internet/NREN: Perspectives, Issues, and Challenges*. He has published articles and reports related to ANSI/NISO Z39.50 and information technology standards development, such as his recent article in ACM's StandardView, "The Government Information Locator Service: A User-Based Approach to Standards." Moen teaches information systems design, information technology standards, and the organization of information. He recently accepted a faculty position at the School of Library and Information Sciences, University of North Texas.

MARK MONMONIER is Professor of Geography, Maxwell School of Citizenship and Public Affairs at Syracuse University. Before joining the geography faculty at Syracuse in 1973, Mr. Monmonier taught at the University of Rhode Island and the State University of New York at Albany. He has served as editor of *The American Cartographer* and president of the American Cartographic Association, and received a Guggenheim Fellowship in 1984. He has published numerous papers on map design, automated map analysis, cartographic generalization, the history of cartography, statistical graphics, geographic demography, and mass communications. His current projects include studies of cartography's role in land use decisions and a book on the uses and design of hazard zone maps.

SHERYL G. OLIVER is GIS Manager at the Illinois Department of Energy and Natural Resources. She is on the Board of Directors of both the National States Geographic Information Council and the Illinois GIS Association. Ms. Oliver is also the Coordinator of the Illinois Geographic Council and a member of the Federal Geographic Data Committee's

CONTRIBUTORS

Framework Working Group to help in the development of a National Digital Geospatial Data Framework.

LINDA C. SMITH is Associate Professor in the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. She joined the faculty in 1977 and teaches in the areas of reference, science reference, and online information systems. Her research interests include information retrieval, artificial intelligence, and science information. She is active in a number of professional associations including the American Society for Information Science, the American Association for the Advancement of Science, and the Association for Library and Information Science Education.

ANNE WATTS is Coordinator of Information and Technology Services at St. Louis Public Library, a position she has held for eighteen years. She also has an appointment at the University of Missouri, School of Library and Information Science, where she instructs extension courses on Bibliography and Reference, Developing Library Collections, and Government Publications in St. Louis. Ms. Watts is chair-elect of the Federal Documents Task Force of the Government Documents Roundtable; Vice-President of the Greater St. Louis Chapter of the Urban and Regional Information Systems Association; and Secretary of the Greater St. Louis Chapter of the American Statistical Association.

INDEX

- AACR2. *See* *Anglo-American Cataloguing Rules*
- Advanced Very High Resolution Radiometer (AVHRR), cataloging of, 22
- Advertising. *See* Publicity
- Aerial survey data: 7-8, 27, 81, 87, 90, 100, 135, 152, 216, 217
- Agricultural Research Service: and the GCDIS, 128
- Alexandria Project, 1, 17-18, 24, 26, 27, 29, 70, 121, 133, 197
- Algorithms for information retrieval, 84, 95-97, 136, 140
- Allen, G. L., 154, 172
- Allen, R. S., 196, 203, 204, 205, 207, 208, 212
- Allison, B., 3, 4
- Allometry: and the growth of digital libraries, 69-80
- American Geological Institute, 150
- American Library Association (ALA), 181, 192
- American National Standards Institute, 2, 19, 47, 66
- Anderson, J. R., 96, 121
- Anglo-American Cataloguing Rules*, 1, 19, 20, 22, 23, 24, 25, 28
- Anonymous FTP. *See* FTP
- ANSI, 2, 19, 47, 66
- Appleyard, D., 153, 171
- ARC/INFO software, 224
- Archaeologic probability map: and the Illinois State Museum, 223
- ArcUSA, 219
- ArcView software, 197, 209, 216, 217, 218, 219, 224
- Armstrong, Mark P., 3, 155, 171
- askSam: and the Electronic Atlas project, 219
- Assessment of Damage and Casualties for Six Cities in the Central United States Resulting from Earthquakes in the New Madrid Seismic Zone*, An, 175, 179
- Assisted Search for Knowledge (ASK) project, 127-128
- Association for Library Collections & Technical Services, 27, 29
- Association of Research Libraries (ARL): and geographic information system research projects, 197, 198
- ASTM Section D18.01.05 on Mapping and GIS, 60
- Atlases, 3, 7, 8-9, 197, 213-219. *See also* Digital maps
- Atmospheric Radiation Measurement Archive: and the GCDIS, 128
- Automated Cartographic Information Center, 3
- Automatic indexing. *See* GIPSY; Indexing
- AVENUE software: and the PLUS+GIS project, 209
- AVHRR (Advanced Very High Resolution Radiometer), cataloging of, 22
- Baldwin, C., 196, 212
- Beard, D., 182, 192
- Benbasat, I., 154, 171
- Bennett, W. D., 187, 192
- Bertalanffy, L. V., 75, 80
- Bertin, J., 155, 171
- Bieger, G. R., 154, 171
- Blakemore, Michael, 14, 15
- Bloom, B. S., 155, 171
- Board of Geographic Names, 142
- Bon, R., 71, 80
- Boolean logic, 84, 139
- Bounding coordinates, 28, 57, 59, 94, 131, 132, 136, 139
- Bourbeau, L., 154, 171
- Bremmer, S. W., 79, 80
- Bretherton, F. P., 19, 29
- Brimicombe, A. J., 88, 121
- Brinker, B., 82, 121
- Browsing. *See* Spatial browsing
- Bureau of the Census. *See* Census Bureau
- Butler, Dixon, 126
- Buttenfield, Barbara P., 2, 69-80
- California Department of Water Resources, 97
- Canadian National Atlas data: and the World Wide Web, 109
- Carbon Dioxide Information Analysis Center: and the GCDIS, 126, 128, 135
- Carter, J. R., 15
- Cartographers. *See* Mapmakers
- Cataloging of geographical information, 1, 17-30, 93, 209. *See also* *Anglo-American Cataloguing Rules*; Indexing; MARC

INDEX

- CCNet, 3
- Census Bureau: and the Electronic Atlas project, 213, 214, 215, 216, 217, 218; and the GCDIS, 128; and the GeoWeb Project, 140-141; and the Illinois Department of Energy and Natural Resources CD-ROM, 222, 223; and market analysis, 187; and TIGER, 103, 105, 106, 109, 120, 177, 184, 196; and the TMS Experimental Browser, 148, 149
- Census market variables, 191
- Census tracts, definition of, 192
- Central Intelligence Agency: and the World Data Bank II data, 114
- Central United States Earthquake Consortium (CUSEC): and emergency scenarios, 175
- Champaign County (IL): and geographic information systems, 3
- Chartrand, K. C., 176, 179
- Chartrand, Robert Lee, 3, 173-179
- Chen, J., 110, 121
- Cheshire II online catalog system, 119
- Chrisman, N., 39
- Christian, E., 44, 46, 66
- Clearinghouse for Networked Information Discovery and Retrieval (CNIDR), 140
- Cohn, J. M., 70, 80
- Collection development: and geographic information, 178
- Collection maintenance: and digital libraries, 70
- Computer equipment: and geographic information system implementation, 196, 200, 201, 202, 203, 205
- Computer skills: and public library users, 204, 205, 207, 208
- ConQuest, 134, 135, 150
- Consistency: and cataloging, 22
- Consortium for International Earth Science Information Network (CIESIN), 129, 142
- Content Standards for Digital Geo-spatial Metadata, 1-2, 20, 24, 25, 27, 28
- Cooperative State Research, Education, and Extension Service: and the GCDIS, 128
- Coordinates. *See* Bounding coordinates
- Coordination of Surveying, Mapping, and Related Spatial Data Activities*, 58
- Coughlin, R. E., 181, 187, 192
- Cruise missiles. *See* Defense technology
- Crustal Dynamics Data Information System (CDDIS): and the GCDIS, 129
- CUSEC (Central United States Earthquake Consortium): and emergency scenarios, 175
- D'Elia, G., 187, 192
- Dalrymple, P. W., 70, 80
- Data compression: and digital libraries, 76-77
- Data Interchange Format (DIF), 136
- Data procurement, 183, 200, 203
- Data quality, 77
- Data retrieval: compared with information retrieval, 83-85
- Data volume: and digital libraries, 69, 75, 76-77
- DataWare Technologies, 139
- De Florian, L., 85, 86, 121
- Decennial Census collection: and the St. Louis Public Library, 214
- Decision support systems: and geographic information, 8, 174-175
- DeCola, L., 71, 80
- Defense Technical Information Center thesaurus, 134
- Defense technology, effect on mapping, 5, 6
- Department of Commerce: and the GCDIS, 128
- Department of Energy: and the GCDIS, 126, 128; and the GCMD, 140
- Department of the Interior: and the GCDIS, 128
- Dervin, B., 159, 171
- Dewey Decimal classification: and digital libraries, 70
- Dexter, A. S., 154, 171
- DiBiase, D., 100, 121
- Digital libraries: and geographical information, 2, 69-80, 81-82, 87, 91-93, 94, 120-121
- Digital library projects, 1, 17, 93, 100-102, 104, 119, 121, 141, 197
- Digital Line Graph format, 140
- Digital maps, 9-10, 91-93, 100-109, 153-155, 156-159, 163-165, 183, 213-219, 221-224. *See also* Atlases; Experiential maps; User interfaces

INDEX

- Dissemination of geospatial data, 32, 33, 35, 36 39, 44, 45, 140-141. *See also* Internet; World Wide Web
- Distance and buffer zone queries, 87, 88-89
- Distributed Active Archive Centers: and the GCDIS, 128, 137
- DLG format, 140
- Dodd, S. A., 19, 30
- Domaratz, Michael, 1, 17, 31-40
- Dumais, S. T., 100, 122
- Dutton, G., 79, 80
- E-systems, 134, 135, 150
- Earth Observation System Data and Information System (EOSDIS): and the Global Change Data and Information Center, 129, 142
- Earth Observing System (EOS), 137
- Earth Resources Observation Systems Data Center, 128-129
- East-West Gateway Coordinating Council: and the St. Louis urban atlas, 214, 219
- Eastman, L. A., 181, 192
- Economic Research Service: and the GCDIS, 128
- EISC, 177
- Electronic Atlas project, 3, 197, 213-219
- Electronic maps. *See* Digital maps
- Electronic storage media: and *Technological Transition in Cartography*, 10
- Elshaw, S., 185, 193
- Emergency Information Systems International: and database of worse-case scenarios, 177
- Emergency Operation Centers (EOC): and special libraries, 177
- Emergency preparedness and response: and geographic information systems, 173-179
- Energy Efficiency and Renewable Energy Network (EREN): and the GCDIS, 128
- Energy Information Administration: and the GCDIS, 128
- Environmental Digital Library, 100-102, 104, 119
- Environmental Protection Agency: and the GCDIS, 128; and the GCMD, 135; and the National Geophysical Data Center, 139
- Environmental Research Laboratory: and the National Geophysical Data Center, 139
- Environmental Systems Research Institute: and geographic information system software, 197, 209, 215, 224
- EOSDIS Version 0 Information Management System, 137-138
- ESRI. *See* Environmental Systems Research Institute
- "Establishment of Government Information Locator Service," 43
- Evansville-Vanderburgh County public library system, 3, 181, 185-189
- "Expanding Research and Development on the ANSI/NISO Z39.50 Search and Retrieval Standard," 46
- Experiential maps, 15. *See also* Digital maps
- Farrell, B., 18, 30
- Farrell, Michael, 127
- Federal Emergency Management Agency (FEMA) maps, 223
- Federal Geographic Data Committee (FGDC), 1, 17, 33, 36, 38, 39, 40, 58, 60, 127, 136, 139, 224
- Federal Information Processing Standards (FIPS) Publication, 45, 192
- Federal Insurance Rate Maps (FIRM), 223
- Field studies: and geospatial data, 152
- File Transfer Protocol, 24, 35
- Fish and Wildlife Service: Hotlist, 142; and the National Geospatial Data Clearinghouse, 36
- Flood zone boundaries data set, 223
- Florida State University: and the PLUS+GIS project, 165, 208, 209
- Forest Service: and the GCDIS, 128
- "Formalizing Cartographic Knowledge," 80
- Forms-based spatial searching interface, 139. *See also* User interfaces
- Frank, A. U., 85, 121
- Free-text searching, 131. *See also* Spatial queries
- Freestone, R., 182, 192
- Frew, J., 86, 121, 122
- FTP, 24, 35
- Furnas, G. W., 100, 122

INDEX

- Galileo, 73, 74
 Gallup Org., 187, 192
 Gardiner, V., 15
 Gatlinburg Symposium on Information Technology and Emergency Management, 176
 GC-ASK Project, 134-135, 142
 GCDIS (Global Change Data and Information System), 125-130, 142
 GCMD (Global Change Master Directory), 129, 135-137, 142
 GENEmap, 135
 General Material Designations (GMDs): and AACR2, 20-21
 General Services Administration: and GILS, 46
 Generic Record Syntax (GRS), 54, 57
 GEO Profile, 60
 Geocoding, 131, 132
 Geodesy and land survey, 7
 Geographic browsing toolkit, 110-119
 Geographic-coordinate indexing. *See* GIPSY; Indexing
 Geographic coordinates. *See* Bounding coordinates
 Geographic information project terms, acronyms of, 144-145; collections of, 96, 134; standardization of, 178. *See also* GIPSY
 Geographic Information Referral Page, 109
 Geographic information retrieval, definition of, 83, 85
 Geographic Information Retrieval and Analysis System (GIRAS), 96, 141
 Geographic information systems (GIS), definition of, 152-153, 195-196; ideal characteristics of, 130-133; implementation model, 202-208; problems with, 183
 "Geographic Information Systems (GISs) and Academic Libraries," 1, 4
 Geographic Names Information System (GNIS), 96, 141
 Geographic queries. *See* Spatial queries
 Geographic Resources Analysis Support System (GRASS), 103, 130, 150
 GEOLIN system, 138-139
 Geometric spatial relationships, 86
 GEONET-L, 142
 GeoRef, 131, 150
 Georeferenced information: compared with planetospatial data, 18
 Georeferenced Information Processing System. *See* GIPSY
 Geospatial data. *See* Spatial data
 Geospatial maps. *See* Digital maps
 Geospatial queries. *See* Spatial queries
 Geospatial text. *See* Spatial text
 GeoWeb Project, 140-141, 142, 148-149
 Getz, M., 181, 182, 187, 192
 Ghosh, A., 192
 GILS: background of, 41-44; and the GCDIS, 127; GILS Attributes Set, 51-52, 54, 55; GILS Core Elements, 50, 59, 60; GILS Locator Records, 41, 43, 44-45, 49-51, 52, 54, 57, 58-59, 60, 61; GILS Profile, 45-47, 50, 52, 55-56, 57, 58, 59, 61; GILS Schema, 52, 54; overview of, 2, 44-45; and spatial information sources, 58-60; and Z39.50, 47-58, 110
 GIPSY, 94, 95-100, 141
 GIRAS (Geographic Information Retrieval and Analysis System), 96, 141
 GIS. *See* Geographic information systems
 "Global Change and the Role of Libraries," 1, 4
 Global Change Assisted Search for Knowledge. *See* GC-ASK Project
 Global change data sets on CD-ROM, 139, 146-147
 Global Change Data and Information System (GCDIS), 125-130, 142
 Global Change Data Management Working Group, 125, 126-128
 Global Change Master Directory (GCMD): and the Global Change Data and Information Center, 129, 135-137, 142
 Global Change Research Information Office, 127, 129, 142
 Global Change Research Program Data Server: and the GCDIS, 128
 Global Ecosystems Database Project, 139
 Glock, M. D., 154, 171
 Gluck, Myke D., 1-4, 58, 66, 151-172
 Goldhor, H., 182, 193
 Goodchild, M. F., 14, 15
 Gore, Albert, Jr., 175, 178
 Gorman, M., 21, 30
 Gould, P., 154, 171
 Gould, S. J., 71, 80

INDEX

- Government agencies: and spatial data, 31-40, 125-150, 174-175, 183
- Government Information Locator Service. *See* GILS
- Graphical interfaces. *See* User interfaces
- GRASS, 103, 130, 150
- GRASSLinks system, 103, 107, 108, 109
- Griffiths, A., 90, 94, 122
- Grundt, L., 181, 192
- GSA: and GILS, 46
- GSNI (Geographic Names Information System), 96, 141
- Guidelines for Bibliographic Description of Reproductions*, 27, 30
- Guidelines for Description of Internet Resources*, 24, 30
- Guidelines for the Preparation of GILS Core Entries*, 51, 66
- Guptill, S. C., 152, 171
- Haire, 71, 74-75, 76, 80
- Hayes, R., 181, 187, 192
- Hayes-Roth, B., 154, 172
- Hazardous Waste Research and Information Center: and digital spatial data on CD-ROM, 221, 223
- Hernon, P., 1, 4
- Hill, Linda L., 2, 125-150
- Hinnebusch, M., 47, 66
- Hollway, J., 15
- Holmes, D. O., 82, 94, 122
- HTML (Hypertext Markup Language) protocol, 109
- Human information processing models, 154
- Huse, Susan, 103
- Hypermaps. *See* Digital maps
- I-Site software, 140
- IBM Corp.: and the overlapping polygon, 178
- Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators*, 46, 66
- Illinois Department of Energy and Natural Resources: and digital spatial data on CD-ROM, 221-224
- Illinois Natural History Survey: and digital spatial data on CD-ROM, 221, 223
- Illinois Office of Research and Planning: and digital spatial data on CD-ROM, 221, 223
- Illinois State Agency Historic Preservation Act, 223
- Illinois State Data Center: and the Electronic Atlas project, 213, 214
- Illinois State Geological Survey: and digital spatial data on CD-ROM, 221, 223
- Illinois State Museum: and digital spatial data on CD-ROM, 221, 223
- Illinois State Water Survey: and digital spatial data on CD-ROM, 221, 223
- Image of the City, The*, 153
- Indexing: of geographic information, 69, 75, 77, 82, 83-84, 178; automatic indexing, 2, 81, 93, 94, 95-100
- Information equivalence: definition of, 154
- Information Infrastructure Task Force, 43, 66
- Information retrieval: compared with data retrieval, 83-85
- Information retrieval algorithms, 84, 95-97, 136, 140
- Information retrieval queries. *See* Spatial queries
- Informativeness measure: definition of, 154; of maps compared with text, 156
- Interagency Working Group on Data Management for Global Change: and GILS, 46
- Interfaces. *See* User interfaces
- International Directory Network (IDN): and the GCMD, 136
- International Organization for Standardization/International Electrotechnical Commission, 46, 66
- International Research Associates, 187, 192
- Internet: cataloging of resources on, 24; and dissemination of spatial data, 32, 33, 35, 36, 44, 45, 140-141; examples of geographic information systems on, 100-109, 142-143. *See also* World Wide Web
- Inventory of Landbased Disposal Sites database, 223
- Ionia Net Browser, 142
- Isometric models, 73
- Java programming language, 110
- Johnson and Johnson: spatial information systems use by, 177
- Johnston, Douglas, 3

INDEX

- Jones, W. P., 100, 122
- Joselyn, Mark, 3, 221-224
- JPEG compression, 76
- Jue, Dean K., 3, 195-212
- Karrow, R. W., Jr., 16
- Keller, C. P., 16
- Kemeny, J. G., 71, 80
- Kofron, C., 197, 212
- Kollen, C., 196, 212
- Koontz, Christie, 3, 181-193
- Kotler, P., 185, 192
- Kronus, C. L., 187, 192
- Lam, N. S., 71, 80
- Lamont, M., 196, 197, 212
- Lan, T. N., 153, 172
- Landsat imagery, 59, 64, 65, 216, 217
- Landview software: and the PLUS+GIS project, 209
- Larkin, J. H., 154, 171
- Larsgaard, Mary Lynette, 1, 17-30
- Larson, Ray R., 2, 81-123
- LASR Project, 142
- Lassen Geographic Browser, 111, 112, 113, 114
- Latitude. *See* Bounding coordinates
- Laurini, R., 86, 91, 122
- Lehrberger, J., 154, 171
- Lentz, R., 165, 171
- Leon County Public Library (Tallahassee, FL): and geographic information projects, 165-166, 198, 209
- Lesgold, A. M., 165, 171
- Levie, W. H., 165, 171
- Levin, J. R., 165, 171
- Liability: and digital data, 224
- Librarians: new roles for, 70, 79, 125, 141, 196
- Library Access, Search, and Retrieval (LASR) Pilot Project, 128
- Library of Congress: and the Alexandria Project, 197
- Library of Congress classification: and geographic information, 28, 70, 82, 94
- Library market analysis. *See* Market analysis
- Library Planning and Decision Making Systems*, 192, 193
- Licensing: and digital data, 224
- Lindsay, P. H., 154, 171
- Lloyd, R., 154, 171
- Long-Term Ecological Research Network: and the Global Change Data and Information Center, 128
- Longitude. *See* Bounding coordinates
- Longstreth, K., 1, 4
- Lutz, M., 1, 4
- Lynch, C. A., 50, 66
- Lynch, K., 153, 171
- Mace, Tom, 127
- MacEachren, Alan, 14, 16
- "Making GIS a Part of Library Service," 1, 4
- "Management of Federal Information Resources," 42, 43
- Mandelbrot, B. B., 71, 80
- Map and Image Laboratory: and Project Alexandria, 197
- Map publishing, 9-10
- Map Resource Center (Tallahassee, FL), 165-166, 209
- Map series, indexing of, 77
- Map use, 153, 157-159, 166-167
- Map widgets, 115-119
- Map-based interfaces. *See* User interfaces
- Mapmakers: amateur, 13; and scale, 69-70
- Mapping policy, 8, 11-12
- Maps. *See* Atlases; Digital maps; Experiential maps
- Maps with the News: The Development of American Journalistic Cartography*, 5, 15, 16
- MARC, 17, 18, 20, 21, 24, 25, 27, 28, 29, 38, 57, 58, 77
- Marc's Favorite Internet Sites, 142
- Marchionini, G., 33, 36, 40
- Marine geological data sets, 138-139, 142
- Market analysis, 3, 181, 182-189, 207-208
- Marketing Institute, 187, 193
- Maurer, H., 33, 36, 40
- McCann, C. M., 100, 122
- McClure, C. R., 42, 43, 46, 66
- McGlamery, P., 196, 197, 212
- McGraw, R., 198, 203, 212
- McLafferty, S. L., 192
- Medyckyj-Scott, D. J., 90, 122
- Meredith, Les, 126
- Metadata, 17, 18, 19, 24-25, 35-39, 43-44, 58-60, 69, 75, 77-78, 135, 136,

INDEX

- 140, 148, 209. *See also* GILS; Project Alexandria; Standards
- Michael, J. J., 47, 66
- Miller, G. A., 96, 122
- Mission to Planet Earth, 137
- Moen, William E., 2, 41-67
- Molz, R. K., 70, 80
- Mondschein, Lawrence, 176, 179
- Monmonier, Mark, 1, 5-16, 153, 172
- Morris, B., 100, 122
- Morton indexing, 77
- Mosaic, 24, 109, 139. *See also* Internet; World Wide Web
- MPEG compression, 76
- Muehrcke, P. C., 155, 162, 172
- Multilevel description: and planeto-spatial data, 25-27
- Multimedia queries, 87, 89-90, 93. *See also* Spatial queries
- Myers, D., 192, 193
- NARA: and GILS, 46, 51, 66
- Naroll, R. S., 75, 80
- NASA, 126, 129, 134, 135-137
- National Academy of Sciences: and the GCDIS, 129
- National Aeronautics and Space Administration. *See* NASA
- National Agricultural Library: and the GCDIS, 128
- National Archives and Record Administration. *See* NARA
- National Biological Service Global Change Research Program Data Management Center: and the GCDIS, 129
- National Center for Atmospheric Research: and the GCDIS, 128; and the GCMD, 136
- National Center for Geographic Information and Analysis, 80, 195
- National Digital Cartographic Data Base—Large Scale, 59, 62
- National Environmental Data Inventory: and the GCDIS, 127
- National Geophysical Data Center, 138-139, 146-148
- National Geospatial Data Clearinghouse, 2, 33-36, 37, 109, 110, 139-140
- National Information Infrastructure, 43
- National Information Standards Organization: and GILS, 47, 52, 66
- National Institute of Standards and Technology, 45, 66
- National Library of the Netherlands, cataloging practices at, 22
- National Map Collection of Canada, cataloging practices at, 25
- National Oceanic and Atmospheric Administration: and the GCDIS, 126, 128, 135, 136, 138-139
- National Performance Review (NPR), 43
- National Research Council, 173, 174, 179
- National Science Foundation: and digital library projects, 1, 17, 128, 197
- National Spatial Data Infrastructure (NSDI), 2, 31-40, 109, 139
- National Technical Information Service: and the GCDIS, 128
- National Wetlands Inventory, 36, 109
- Natural language queries, 135. *See also* Spatial queries
- Natural resources data sets, 221-224
- Nebert, D. D., 60, 109, 110, 122, 133, 140, 150
- Negative scribing, 10
- Netscape, 109. *See also* Internet; World Wide Web
- New York, hypermaps of, 103, 105, 106
- Newby, G. B., 160, 162, 172
- Newman, I. A., 90, 122
- Nielson, B., 79, 80
- NII, 43
- Nilan, M. S., 159, 160, 171
- NISO: and GILS, 47, 52, 66
- NOAA. *See* National Oceanic and Atmospheric Administration
- Norman, D. A., 154, 171
- Noronha, V. T., 14, 16
- NTIS: and the GCDIS, 128
- Oak Ridge National Laboratory: and the GCDIS, 128
- OASIS (Open Architecture for Scientific Information Systems), 134, 135
- Office of Management and Budget: Bulletin 43, 51, 58, 61, 66, 67, 95-101; Circular A-16, 58; Circular A-42, 43, 130; and GILS, 46
- Oliver, Sheryl G., 3, 221-224

INDEX

- Ousterhout, J. K., 112, 122
- Overlapping polygons, 96-99, 178
- Palmer, E. S., 181, 182, 187, 192, 193
- Parker, V., 25, 27, 30
- Path queries, 87, 89. *See also* Spatial queries
- Patton, G., 24, 30
- Peano curves, 77. *See also* Indexing
- Personnel issues, 195, 199, 200, 201, 202, 203, 205, 209, 218
- Physical description: and digital data, 20-24
- Planetospatial data, cataloging of, 1, 17-30
- Plaunt, Christian, 94, 123, 141, 150
- Plewe, Brandon, 141, 148, 149, 150
- PLUS+GIS project, 195, 208-210
- Point-in-polygon queries, 87, 93, 109. *See also* Spatial queries
- Precision, 130-131
- Principle of Similitude, 73
- Privatization of multinationals, effect on mapping, 11-12
- Procedural knowledge: and map use, 153
- Project Alexandria, 1, 17-18, 24, 26, 27, 29, 79, 121, 133, 197
- Public libraries: and geographic information, 3, 151-152, 167-169, 195-212
- Public Library Association, 182, 183, 189, 193
- Public Library Users Geographic Information System, 195, 208-210
- Public School Library Society of St. Louis, 213
- Publicity: and geographic information system implementation, 202, 203, 213
- Quadtrees, 77. *See also* Indexing
- Rand, R. Y., 1, 4, 127
- Ranking, 84, 132
- Recall, 130-131
- Record Reference Book libraries, 139
- Region queries, 87-88. *See also* Spatial queries
- Regional Commerce and Growth Association (St. Louis): and digital maps, 216, 218
- Regional Research and Development Services: and the Electronic Atlas project, 214, 215, 216
- Reid, H., 186, 193
- Relevance, 130
- Remote sensing data, 7-8, 81, 90, 100, 152. *See also* Aerial survey data
- Representative Fraction (RF), 71-72
- Richardson, L. F., 71, 76, 80
- "Role of Spatial Information Systems in Environmental Emergency Management, The," 176-177, 179
- Royal Library (Netherlands), cataloging practices at, 22
- Ruggles, C. L. N., 90, 122
- Rural market areas: and public libraries, 182
- Rushton, Gerard, 3
- Ryan, J., 43, 66
- Sales, Georgia, 178, 179
- Salton, G., 84, 85, 122
- San Francisco Bay area, hypermaps of, 100-103, 107, 108, 109, 120
- Satellite data. *See* Aerial survey data
- Scale, 29, 69-76, 78, 152-153
- Schamber, L., 162, 172
- Schlipf, F. A., 181, 182, 187, 193
- Schroeder, R. Christopher, 3
- SDTS format, 140
- Search queries. *See* Spatial queries
- Security, 12, 200
- Semantic network retrieval engines, 134, 150
- Sequoia 2000 project, 110
- Shaughnessy, T. W., 181, 193
- Shoham, S., 182, 193
- Siegel, M. A., 70, 80
- Simon, H., 154, 171
- Simple Unstructured Text Record Syntax, 57-58
- Singley, P. T., 19, 29
- Slocum, T. A., 14, 16
- Smith, B. W., 187, 192
- Smith, Linda C., 1-4, 70, 80
- Smits, J., 22
- Snyder, J. P., 16
- Socioeconomic data, 32, 184, 186
- Sol Katz List of Spatial Geographic Search Sites, 143
- Southern Illinois University at Edwardsville: and the Electronic Atlas project, 213, 214
- Spatial Analysis, Research, and Training (SART) Program: and the GIS+PLUS project, 209

INDEX

- Spatial browsing, 2, 82, 91-93, 100-109, 110-119, 133, 138, 139, 143, 149
- Spatial data: characteristics of, 58; compared with geospatial data, 18, 139, 152; compared with planetospatial data, 18; definition of, 18, 32; and digital library growth, 76; dissemination of, 32, 33, 35, 36, 39, 44, 45, 140-141; and FTP, 35; maps vs. text, 155-159; potential uses, 173-174; use studies, 155-169, 202-203; user community characteristics, 31-32, 204, 207
- Spatial Data Transfer Standard format, 140
- Spatial metadata. *See* Metadata
- Spatial queries, 81, 85-89, 93, 94, 100, 109, 131-133, 155, 162-163
- Spatial text, 153-154, 155-159, 163-165, 177
- Spatial WAIS, 137, 139, 148
- Special libraries: and emergency preparedness and response, 3, 173-179
- SPOT Imagery, 216, 217
- St. Louis Public Library: and the *St. Louis Public Library Electronic Atlas*, 3, 197, 213-219
- Staff issues. *See* Personnel issues
- Standards, 2, 9; for emergency management databases, 176; for geographic information terminology, 178; for metadata, 17, 24-25, 35-39, 136, 139-140, 209, 224. *See also* *Anglo-American Cataloguing Rules*; MARC; X3.27; Z39.50
- Stasz, C., 153, 154, 172
- State University of New York at Buffalo: and the GeoWeb Project, 141, 148; and Project Alexandria, 197
- Steinke, T., 154, 171
- Stonebraker, M., 110, 122
- Strategies and Systems for Disaster Survival*, 176, 179
- Streeter, L. A., 154, 172
- Stuber, Chris, 141, 148
- SUTRS, 57, 58
- Syracuse University: and GILS, 46
- "Taxonomy of Human Services, A," 178, 179
- Tcl/Tk programming language, 112-119
- Technological Transition in Cartography*: author's conclusions, 10-13; critical reviews, 13-15; organization and content, 6-10; overview, 1, 5-6
- Terrestrial Ecosystems Regional Research and Analysis (TERRA) Laboratory: and the GCDIS, 128
- Text. *See* Spatial text
- Thesauri of geographic information terms: and the GIPSY project, 95, 96; need for, 178
- Thompson, D., 86, 91, 122
- Thompson, D. W., 73, 80
- Thorndyke, P. W., 153, 154, 172
- Thrall, G. I., 185, 193
- TIGER, 103, 105, 106, 109, 114, 120, 177, 184, 196, 216
- TigerWWW, 149
- Tillett, Barbara, 27, 30
- Time: and metadata, 28
- Tioga system, 110-111
- TMS Experimental Browser, 143, 149
- Toolkit for geographic browsing, 110-119
- Topological spatial relationships, 86
- Topologically Integrated Geographic Encoding and Referencing. *See* TIGER
- Toward a Coordinated Spatial Data Infrastructure for the Nation*, 173, 179
- Training, 195, 207, 209
- U.S. Bureau of the Census. *See* Census Bureau
- U.S. Department of Commerce. *See* Department of Commerce
- U.S. Department of Energy. *See* Department of Energy
- U.S. Department of the Interior. *See* Department of the Interior
- U.S. Environmental Protection Agency. *See* Environmental Protection Agency
- U.S. General Services Administration. *See* General Services Administration
- U.S. Geological Survey: and county atlases, 7; and the GCMD, 135, 136; and the Geographic Information Referral Page, 109; and the Geographic Information Retrieval and Analysis System (GIRAS), 96; and the Geographic Names Information

INDEX

- System (GNIS), 96; and GILS, 46, 59, 60, 62, 64; and the Illinois Department of Energy and Natural Resources CD-ROM, 3, 222, 223; and the National Geospatial Data Clearinghouse, 36, 109, 139-140; and Project Alexandria, 197
- U.S. Global Change Data and Information System Implementation Plan*, 126, 150
- U.S. National Aeronautics and Space Administration. *See* NASA
- U.S. National Archives and Record Administration. *See* NARA
- Uniform Resource Identifier (URI): and GILS, 44
- Uniform Resources Locator (URL): and GILS, 60
- United Medical Language System (UMLS), 134
- United States Global Change Research Program (USGCRP), goals of, 125
- Universal Transverse Mercator, 119
- University of California at Berkeley: and digital library projects, 94, 100-102, 103, 104, 107, 108, 109, 119, 121, 141
- University of California at Santa Barbara: and Project Alexandria, 121, 133, 197
- University of Minnesota: John R. Borchert Map Library, 3
- University of Virginia: and the Library Access, Search, and Retrieval Pilot Project, 128
- Urban Atlas, Tract Data for Standard Metropolitan Statistical Areas: St. Louis, Missouri—Illinois*, 214
- Urban market areas: and public libraries, 182
- User interfaces: 70, 77, 81, 82, 92, 100, 111, 113, 118, 126, 134, 135, 136-137, 138, 139, 140, 209, 216, 217.
See also Digital maps
- USMARC. *See* MARC
- van Hoff, A., 110, 122
- Van House, N. A., 181, 182, 187, 193
- Vanetti, E. J., 154, 172
- Vaziri, M., 153, 172
- Vector space information retrieval model, 85
- Virtual Tourist system, 103
- Vitello, D., 154, 172
- WAIS, 110, 137, 139, 148
- Walker, A. S., 16
- Walker, D. R. F., 90, 122
- Watts, Anne, 3, 197, 212, 213-219
- Wavelet decomposition, 76-77
- Wax engraving: and mapmaking, 9-10
- Weibel, S., 70, 80
- Wells coverage: and the Illinois Department of Energy and Natural Resources CD-ROM, 223
- Westcott, B., 207, 212
- Wheeler, J. L., 182, 193
- White, R., 154, 171
- Wilensky, R., 33, 40, 100, 119, 123
- Winkler, P. W., 21, 30
- Woldenberg, M. J., 71, 80
- Wong, M., 196, 202, 205, 207, 212
- Wonsiewicz, S. A., 154, 172
- Woodruff, A. G., 94, 123, 141, 150
- WordNet thesaurus, 96
- World Agricultural Outlook Board (WAOB): and the GCDIS, 128
- World Data Bank II data, 114
- World Data Center (WDC), 138
- World Wide Web (WWW), 35, 39, 59-60, 91, 100-110, 136, 137. *See also* Internet
- X Windows-based geographic interfaces, 111, 113, 118, 136, 138, 139.
See also User interfaces
- X3.27, 19. *See also* Standards
- Z39.50, 2, 35, 41, 42, 45, 46-58, 60, 110, 127, 134, 150. *See also* Standards
- ZIG (Z39.46 Implementors Group), 47
- Zweizig, D. L., 187, 193

UNIVERSITY OF ILLINOIS-URBANA



3 0112 077017751